THE USE OF MAIZE STREAK VIRUS (MSV) REPLICATION-ASSOCIATED PROTEIN MUTANTS IN THE DEVELOPMENT OF MSV-RESISTANT PLANTS

Dionne Natalie Shepherd

Thesis presented for the degree of Doctor of Philosophy in the Department of Molecular and Cell Biology,
University of Cape Town
February, 2003

Certification by Supervisors

In terms of paragraph GP9 of the regulations for the degree of Doctor of Philosophy at the University of Cape Town, I certify that I approve of the inclusion in this thesis of material already published, or submitted for publication by candidate Dionne Natalie Shepherd.

Professor J. A. Thomson
Department of Molecular and Cell Biology
University of Cape Town

Department of Molecular and Cell Biology
University of Cape Town

February, 2003

Abstract

Maize streak virus (MSV) is the type member of the *Mastrevirus* genus of the *Geminiviridae*. As the causal agent of maize streak disease (MSD), MSV is the most significant pathogen of maize in Africa, resulting in crop yield losses of up to 100%. Transmitted by leafhoppers (*Cicadulina* spp.), MSV is indigenous to Africa and neighbouring Indian Ocean Islands. Despite maize being a crucial staple food crop in Africa, the average maize yield per hectare in Africa is the lowest in the world, leading to food shortages and famine. A major contributing factor to these low yields is MSD.

To genetically engineer MSV-resistant maize using the pathogen-derived resistance (PDR) strategy, the viral replication-associated (Rep) protein gene was targeted, whose multifunctional products Rep and RepA are the only viral proteins essential for replication. *Rep* constructs had previously been made containing deleterious mutations in several conserved amino acid motifs. In this study, these mutants and the wild type Rep gene were truncated to remove key motifs involved in viral replication. A quantitative PCR assay was developed to determine the effects of the mutant and truncated Reps on viral replication in black Mexican sweetcorn (BMS) suspension cells. The MSV-sensitive grass *Digitaria sanguinalis* was then transformed with *Rep* constructs that inhibited MSV replication in BMS, and transgenic lines were tested for virus resistance. Several plants of a *D. sanguinalis* line transgenic for a mutated full-length Rep gene showed excellent resistance (immunity) to MSV, but the transgene had negative effects on aspects of plant growth and development. Transformation with a mutated/truncated Rep gene resulted in healthy fertile transgenic *D. sanguinalis* plants, many of which showed good MSV resistance. Fertile maize (Hi-II) T_1 transgenic plants expressing the truncated/mutated Rep gene have been obtained, the offspring of which will be tested for resistance to MSV. Considering the success in achieving MSV-resistant *D. sanguinalis*, there is good reason to believe that the transgenic maize will too be resistant to MSV.

The transient expression studies in BMS provided some interesting insights into the mechanics of MSV replication and its interaction with host factors. A Rep construct with a mutation (Rb^-) abolishing the protein's interaction with the host retinoblastoma-related (RBR) protein, previously thought to be required for viral replication, surprisingly supported high-level viral replication in BMS. A virus carrying the Rb^- mutation was infectious in maize; however, one of the nucleotides of the three-nucleotide mutation reverted at an extremely high frequency. A study was carried out to

determine the time taken for the nucleotide reversion to occur, and the point at which the revertant population superseded the original mutant population. These data will possibly enable the mutation rate of MSV to be calculated, which is valuable information when attempting to design resistance strategies that cannot be overcome by mutation of the viral genome. Further studies were aimed towards determining the selective advantage and the absolute requirement of the single nucleotide reversion for viral infectivity in maize, as well as its effect on viral replication in BMS.

In an attempt to further define the role of RepA in the virus' lifecycle, separate intronless Rep and RepA constructs were made, both wild type and containing the Rb^- mutation. The effects of the presence and absence of RepA on viral replication and infectivity in maize were determined. It was concluded that the Rb^- mutation had no effect on the role of Rep or RepA in the initiation of viral replication in BMS suspension cells. Overexpression of RepA inhibited the replication functions of Rep in BMS, and the absence of RepA had the effect of slightly reducing replication levels. These data indicate that a precise balance of Rep and RepA is required for optimal replication of the viral genome, and altering the wild type ratios of the two proteins has a negative effect on viral replication. RepA was required for efficient infectivity of MSV in maize, although results indicated that a mutant virus unable to express RepA may be capable of establishing a very weak infection.

Acknowledgements

It is difficult to know where to begin: there are so many people who have contributed to this project.

A good starting point would be to thank my previous lecturer and supervisor at Rhodes University, **Prof. Don Hendry**. This wonderful man was responsible for kindling my interest in virology, so that when it came to choosing a PhD project at UCT, working on *Maize streak virus* was an obvious choice.

From then on it was a pleasure to work with my co-supervisors, **Prof. Jennifer Thomson** and **Prof. Ed Rybicki**, who provided a wonderful mixture of leaving me free to pursue my own day-to-day goals in my own weird hours, but with constant guidance, support and enthusiasm. Although I was reluctant at the time, I must thank Jennifer for making me down tools and start writing this thesis. If it wasn't for her I would still be in the lab trying to "tie up loose ends"! Thanks also to Jennifer and Ed for their generous financial support throughout this project, and for their invaluable constructive criticism of this thesis.

The work in this project could never have been achieved without the help of **Marion Bezuidenhout**, who was more than simply my tissue culture technician, she was indispensable in my life! Only a week before the printing of this thesis Marion was still helping me by extracting DNA from maize leaves. Although she was also the technician of another lab, she never hesitated in helping me, from subculturing the enormous amount of tissue culture involved in this project, to looking after the transgenic plants (her "babies") and helping with hundreds of DNA extractions. Finally, she is a very warm friend. Thank you Marion.

Darren Martin played an enormous role in this project. He was a hands-on supervisor, someone who was always ready to give advice, and responsible for many of the ideas that led to the work in this thesis. I am also extremely grateful to Darren for proofreading the entire Literature Review – an odious task – and for his constructive criticism of many of the results presented in this thesis.

There are many people who contributed towards the work presented in this thesis. Those to be thanked are **David McGivern** for the yeast-two-hybrid analysis; **Rikus Kloppers** for looking after and self-pollinating the maize; **Ziyaad Valley-Omar** for the doing the RNA extractions, **Eric van der Walt** for the isolation of the viral genomes from leafhopper-inoculated maize and doing the sequence alignments; **Di James** for doing all the sequencing reactions in this project, and **Darren Martin** for performing most of the agroinoculations and symptom analysis.

Drs **Margaret Boulton** and **David McGivern** are to be thanked for their help and friendship when I worked in their lab in Norwich, UK, and for their personal communications of various results presented in the Literature Review.

For their friendship, entertainment and assistance over the years I thank my co-workers **Janet Willment**, **Leigh Berrie**, **Arvind Varsani**, **Eric van der Walt** and **Darren Martin**. Janet, I also thank you for taking me under your wing when I first joined the lab and always having time to help and encourage me. I hope that I can play a similar role in a young, unsure, unconfident student's life.

Thank you to the departmental assistants for helping with the day-to-day running of the lab and plant growth rooms, especially **Sheila Koeries**, **Jackie McCann**, **Donald Solomons** and **Melvin Keating**.

Finally, a big thank you to my family: **Mom**, **Geoff**, **Dad** and **Suzie** for all being the most caring and supportive parents I could hope for (you must I have thought I would remain a student forever!); **Carla** and **Dave** for apparently being interested and impressed by my work, and **Ryan** and **Georgia** for just being cute!

The biggest thank you of all to **Richard**, who has been a pillar of support throughout the writing of this thesis and indeed the whole project. He has worn many hats – financer, driver, friend and lover. What would I do without you!

Contents

Chapter 1

Literature review

Contents

1.1 INTRODUCTION

"*The disorder of the mealie plant, locally described as 'Mealie Blight', 'Mealie Yellows' or 'Striped Leaf Disease', belongs to a group of plant troubles arising from obscure causes...*" was how maize streak disease (MSD) was first described by Fuller (1901) in Natal, South Africa. Fuller mistakenly attributed the disease to a soil disorder, but in retrospect it is quite clear that the "mealie variegation" he described and drew in minute detail can be attributed to *Maize streak virus* (MSV). In the hundred years since this first report, scientists have come a long way in identifying and analysing the causal agent of MSD, to the point where we can now design effective strategies to control or even eliminate the disease in maize.

The first milestone in MSD research was reached in 1924, when H.H Storey determined that a virus obligately transmitted by leafhopper species of the genus *Cicadulina* was the causal agent of MSD (Storey, 1924). Storey named the virus *Maize streak virus*. Storey (1931) was also the first to describe the genetic basis of transmission of MSV by *Cicadulina mbila*, and that resistance to MSD in maize could be inherited (Storey and Howland,1967).

MSV particles were first purified and visualised by Bock *et al.* (1974): they were found to have a novel twinned quasi-icosahedral (geminate) shape, from which the name 'geminivirus' was born. This was followed by the unexpected discovery in 1977 that geminivirus particles contain circular single-stranded DNA (ssDNA), a genome type never before observed in plant viruses (Goodman *et al.*, 1977a, b; Harrison *et al.*, 1977). These novel characteristics led to the proposal of a new virus group - the geminiviruses - consisting of MSV and other viruses with geminate particle morphology and ssDNA genomes. This was officially accepted by the International Committee for Virus Taxonomy in 1978. The group was subsequently given the status of a taxonomic family in 1995, and by 2000 the family *Geminiviridae* had three genera (Briddon and Markham, 1995; Rybicki *et al.*, 2000), with MSV as the type member of the genus *Mastrevirus*.

Following the visualisation of circular ssDNA-containing geminivirus particles in the 1970s, the next major advance in geminivirus research came in the early 1980s with the cloning and sequence analysis of the first geminivirus genomes. The characterization and sequencing of the bipartite genome of *African cassava mosaic virus* (ACMV) in 1983 (Stanley and Gay, 1983) was followed by monopartite MSV in 1984 (Mullineaux *et. al.,* 1984). This important development

led to a new age of geminivirus research, with the start of intensive investigation of geminivirus molecular biology.

1.2 THE MOLECULAR BIOLOGY OF GEMINIVIRUSES

The majority of this section will focus on the wealth of information on geminivirus molecular biology obtained in the 20 years since the first geminivirus genome was sequenced, with particular emphasis on MSV. However, in cases where there is little or no information on certain aspects of MSV biology, analogies will be drawn from more intensively studied geminiviruses. The taxonomy and general properties of the different geminivirus genera are summarised below as an introduction to the family.

1.2.1 The *Geminiviridae*

The taxonomic family *Geminiviridae* is classified into four genera (*Mastrevirus*, *Begomovirus*, *Curtovirus* and *Topocuvirus*; see Table 1 and Figure 1) on the basis of their host range, genome organisation and vector specificity. In addition to these biological characteristics, there is also substantial phylogenetic support for the existence of these genera (Rybicki, 1994; Padidam *et al.*, 1995). Mastreviruses (type member *Maize streak virus*) have monopartite genomes and are transmitted by different leafhopper species (Homoptera: family *Cicadellidae*) and generally infect monocotyledonous plants. Begomoviruses (type member *Bean golden mosaic virus*; BGMV), which comprise the largest genus of the family, are transmitted by a single whitefly species (*Bemisia tabaci*), and all infect dicotyledonous plants. Most have bipartite genomes (called DNA A and DNA B), although there are some viruses in this genus that have monopartite genomes (Rybicki *et al.*, 2000). Curtoviruses (type member *Beet curly top virus*; BCTV) occupy an intermediate position between mastreviruses and begomoviruses in that they have monopartite genomes and are transmitted by leafhoppers, but infect only dicotyledonous hosts. The latest addition to the family, topocuviruses (type member *Tomato pseudo curly top virus*) are similar to curtoviruses in genomic organisation but are transmitted by treehoppers.

TABLE 1.1 Classification and general properties of the *Geminiviridae*

Genus	No. of approved members	Examples of members (Type species first)	Genome size (nt)/ Arrangement	Host Range	Vector
Mastrevirus	12[a]	*Maize streak virus* (MSV)	2684-2701/ monopartite	Poaceae (cereals, grasses)	Leafhopper
		Wheat dwarf virus (WDV)	2749/ monopartite	Poaceae	Leafhopper
		Sugarcane streak virus (SSV)	2706/ monopartite	Poaceae	Leafhopper
		Bean yellow dwarf virus (BeYDV)	2561/ monopartite	Leguminosae	Leafhopper
		Tobacco yellow dwarf virus (TYDV)	2580/ monopartite	Solanaceae	Leafhopper
Curtovirus	4[b]	*Beet curly top virus* (BCTV)	2933/ monopartite	Dicot plants	Leafhopper
		Horseradish curly top virus (HrCTV)	3080/ monopartite	Horseradish, Shepherd's purse	Leafhopper
Topocuvirus	1[c]	*Tomato pseudo-curly top virus* (TPCTV)	2861/ monopartite	Solanaceae	Treehopper
Begomovirus	103[b]	*Bean golden mosaic virus* (BGMV)	A: 2644; B: 2609/ bipartite	Leguminosae	Whitefly
		Tomato golden mosaic virus (TGMV)	A: 2588; B: 2522/ bipartite	Solanaceae	Whitefly
		African cassava mosaic virus (ACMV)	A: 2779; B: 2724/ bipartite	Euphorbiaceae	Whitefly
		Tomato leaf curl virus (ToLCV)	2766 / monopartite	Solanaceae	Whitefly
		Squash leaf curl virus (SqLCV)	A: 2634; B: 2606/ bipartite	Cucurbitaceae	Whitefly
		Tomato yellow leaf curl virus (TYLCV)	2743-2790/ monopartite	Solanaceae	Whitefly

[a]Boulton and Davies (2002)

[b]Fauquet *et al.* (2003)

[c]Briddon (2002)

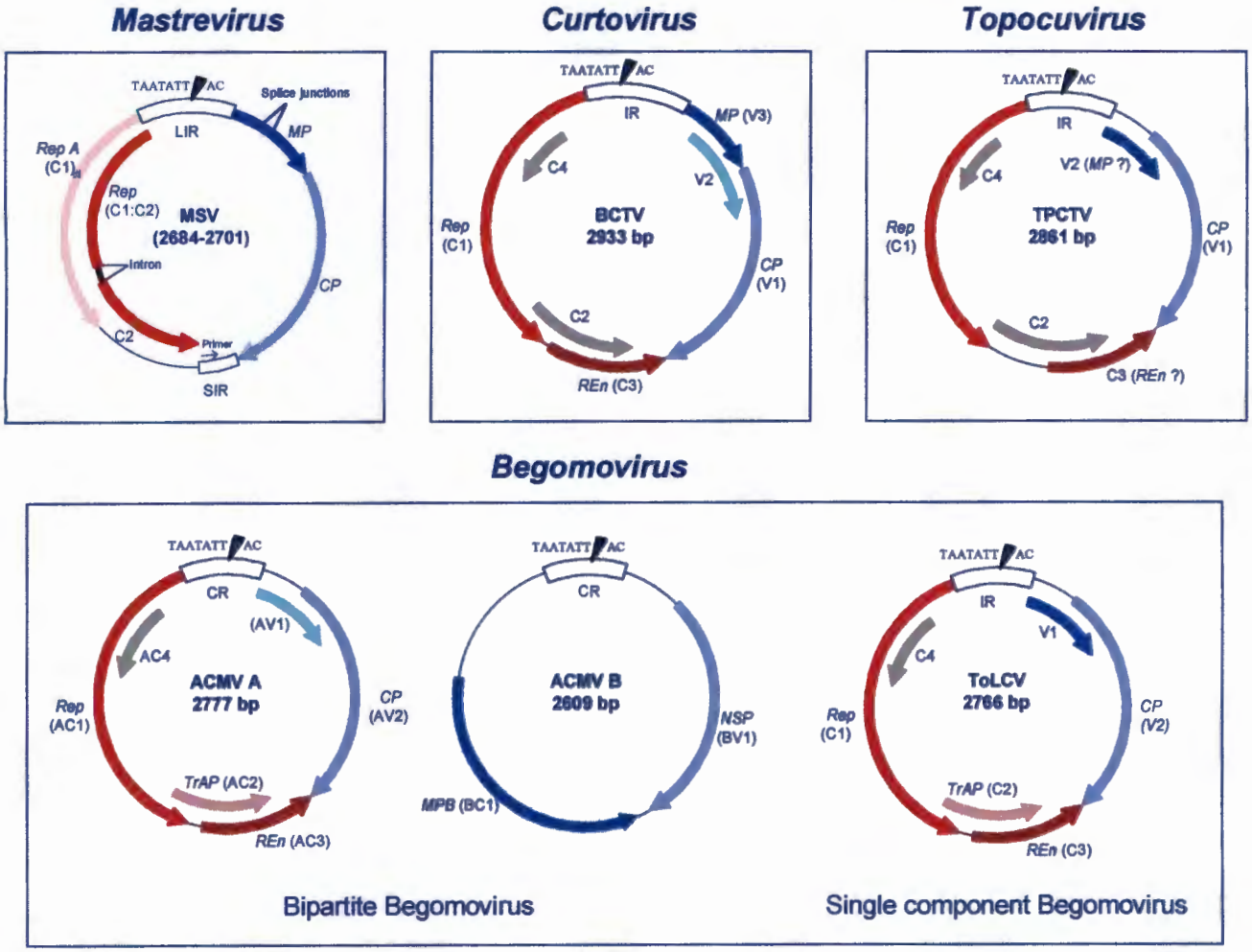

FIGURE 1.1 Genome organization of representative members of the four geminivirus genera. Curved arrows indicate open reading frames, diverging in the complementary (C) and virion (V) senses from an intergenic region (IR) in curtoviruses and topocuviruses, a long intergenic region (LIR) in mastreviruses, and a common region (CR) in bipartite begomoviruses. The CR is a ~200 bp sequence in the begomoviral IR that is conserved between the A and B components of a bipartite begomovirus. The position of the plus strand rolling circle replication (RCR) initiation site (TAATATTAC), situated within the loop of a stem-loop structure, is indicated within each genus' IR. To initiate RCR, a nick (▼) is introduced by the replication associated protein (Rep) at the penultimate A nucleotide of the invariant nonanucleotide sequence. In mastreviruses, bidirectional transcription initiates in the LIR and terminates mainly in the short intergenic region (SIR), which contains signals for polyadenylation. The SIR also functions as the C sense (negative strand) origin of replication. A small (~ 80 nt) ssDNA molecule annealed in the SIR to the plus strand is thought to act as the primer for negative strand replication. Genes have been named according to either their function, if known, or their genetic location. Genes in shades of red have a function in the early stage of the virus life cycle (DNA replication, regulation of transcription, and most likely, interfering with cellular process needed for the replicative cycle). Genes in shades of blue in general have movement and structural functions. Where the function of an ORF is not known, it is coloured in gey. Rep is found in all geminiviruses, although in mastreviruses a unique variant of Rep may also be expressed, called RepA. In this genus, Rep (C1:C2) is expressed from a spliced transcript of ORFs C1 and C2, whereas RepA (C1) is potentially expressed from the unspliced transcript. Other genes with known functions are *CP* (coat protein gene found in all geminiviruses), *MP* (movement protein gene found in mastreviruses and curtoviruses), *TrAP* (transcription activator protein gene in begomoviruses); *REn* (replication enhancer gene found in begomoviruses and curtoviruses); and *MPB* and *NSP* (movement protein gene and nuclear shuttle protein gene, respectively, found on the B component of bipartite begomoviruses). The C4 ORF appears to have different functions in different genera. In curtoviruses it may be involved in tumour induction (Latham *et al.*, 1997), while in monopartite begomoviruses such as TYLCV it is apparently involved in movement (Jupin *et al.*, 1994). Two contrasting studies in the bipartite begomovirus, TGMV, have indicated either a role for C4 in transcriptional regulation of Rep (Groning *et al.*, 1994) or no function at all (Pooma and Petty, 1996). The function of C4 in topocuviruses is unknown. The AV1 ORF in bipartite begomoviruses, sometimes called the pre coat protein gene, or *PreCP*, and the V2 ORF in curtoviruses, may both be involved in regulation of ssDNA accumulation. The V2 ORF in topocuviruses may have a function in movement, but this has not been proven. The topocuvirus C3 gene has substantial sequence homology to REn, but its function is unknown. Similarly, although the C2 gene of curtoviruses has homology to TrAP, its function in curtoviruses is unknown. The role of C2 in topocuviruses has not been determined.

It is generally accepted that geminiviruses replicate in the nucleus using a rolling circle replication (RCR) mechanism (Saunders *et. al.*, 1991; Stenger *et. al.,* 1991). Convincing evidence for this model is the production of supercoiled, open circular, and linear double-stranded (ds) DNA species, a hallmark of the RCR strategy employed by circular ssDNA bacteriophages (Gruss and Ehrlich, 1989; Kornberg and Baker, 1992) and a class of eubacterial plasmids (Baas and Jansz, 1988). As additional evidence for the RCR model, the N-terminal portions of all geminiviral replication associated proteins (Reps) contain three motifs conserved among the replication initiator proteins from other known RC replicons (Koonin and Ilyina, 1992). More recently, electron microscopic visualisation and two-dimensional gel analysis of *Abutilon mosaic virus* (AbMV) confirmed that geminiviruses replicate via rolling circles (Jeske *et al.*, 2001); however, there is a strong possibility that a recombination-related process (recombination-dependent replication; RDR) is an additional replication strategy employed by geminiviruses (Jeske *et al.*, 2001).

As can be seen in Figure 1.1, all geminivirus genomes contain an intergenic region (IR) from which viral genes diverge in both the virion (V) and complementary (C) senses. The IR contains RNA-polymerase II-type promoter sequences that drive the transcription of genes in both genome senses (bidirectional transcription). In general, genes encoded in the virion sense ("V" genes) have functions in virus movement and encapsidation ("late" functions), whereas genes encoded in the complementary sense ("C" genes) are involved in virus replication and transactivation of the virion-sense promoter ("early" functions). The IR of all geminiviruses contains a stem-loop structure within which an invariant nonanucleotide sequence (TAATATT↓AC) contains the initiation site (↓) of RCR (Heyraud-Nitschke *et al.*, 1995; Laufs *et al.*, 1995a; Stanley, 1995).

The only genes shared by all geminiviruses are the Rep and coat protein (CP) genes (Rybicki, 1994; Rybicki *et. al.*, 2000), the remainder of each genus' genes differing both in number and arrangement. Mastreviruses have a unique genome composition (reviewed in detail in the next section), being the only geminiviruses to possess a small, (or short) intergenic region (SIR), a Rep transcript with an intron which, depending on whether or not it is spliced, is capable of expressing a RepA protein (unspliced) or a full-length Rep protein (spliced); and virion sense transcripts that are also spliced, encoding the movement protein (MP) and the CP.

The only geminiviruses with two genome components are begomoviruses, such as the Old World (OW) begomovirus, ACMV, and the New World (NW) begomovirus, BGMV. The C sense

strand of the A component of all bipartite begomoviruses encodes Rep, a transcription activator protein (TrAP), and a replication enhancer protein (REn), while the CP is expressed from the A component's V sense strand. In addition, the OW begomovirus A component contains a C4 gene (AC4) on the C sense strand, whose function is unknown, and a "pre-coat protein" gene (PreCP) on the V sense stand. The latter is apparently involved in ssDNA accumulation (Wartig *et al.*, 1997) and/ or movement (Padidam *et al.*, 1996). The B component encodes two proteins involved in viral movement: on the C sense BC1, or otherwise known as the movement protein (MPB), and on the V sense BV1, or nuclear shuttle protein (NSP). The NSP, although unique to bipartite begomoviruses, shows some sequence homology to geminivirus CPs (Rybicki, 1994). Monopartite begomoviruses such as ToLCV have an arrangement of C sense genes identical to that on the A component of bipartite begomoviruses, while the virion strand contains two overlapping coding regions (V1 and V2), both of which are required for infectivity but not for replication (Rigden *et al.*, 1993; Padidam *et al.*, 1996). Satellite DNAs are frequently associated with monopartite begomovirus infections (Dry *et al.*, 1997; Monsoor *et al.*, 1999; Saunders and Stanley, 1999). Whereas the satellite DNAs appear to play no essential role in the disease caused by their associated begomovirus, a recently detected DNA (named DNA β) associated with the otherwise weakly virulent monopartite begomovirus, *Ageratum yellow vein virus* (AYVV) massively enhances the virus' virulence (Saunders *et al.*, 2000).

Curtoviruses, exemplified by BCTV, have an organization of C sense genes similar to that of begomoviruses, while the V sense organisation is similar to that of the monopartite geminiviruses. This has prompted speculation that curtoviruses arose from a recombination event between mastreviruses and begomoviruses (Rybicki, 1994), although more recent evidence suggests that the recombination event involved the 5' portion of *Rep* only, resulting in curtoviruses having a begomovirus-like Rep (Padidam *et al.*, 1999a; Martin *et al.*, unpublished). BCTV has three virion-sense ORFs: these are V1 (CP), V2 and V3 (MP). V1 and V3 are required for infection but not for replication, suggesting that they both possibly have a role in movement (Briddon *et al.*, 1989), while V2 may modulate the conversion of dsDNA to ssDNA (Hormuzdi and Bisaro, 1993).

The topocuvirus TPCTV has a genomic organization very similar to that of curtoviruses, although the functions of some of its genes have yet to be elucidated. Although the C sense strand has four ORFs (C1, C2, C3 and C4), only the product of the C1 gene (Rep) is known. While the C3 gene encodes a protein with substantial sequence homology to the begomovirus and curtovirus REn, its

function in topocuviruses has not been determined. The topocuvirus V sense strand encodes two potential proteins, V1 (CP) and V2. The latter most likely functions as a MP, although this has not been proven.

1.2.1.1 The evolutionary origin of geminiviruses

There is some similarity between the plant-infecting geminiviruses and nanoviruses, and the vertebrate-infecting circoviruses: all groups have circular ssDNA genomes that multiply by RCR, with each genome component having a stem-loop structure containing a very similar nonanucleotide sequence in the loop. However, nanovirus genomes consist of at least six circular ssDNAs, all about 1 kb in size (Aronson *et al.*, 2000). There is also detectable sequence homology between the Rep proteins of each group, suggesting a similar evolutionary origin. Because of the similarity between these Reps and the replication initiator proteins of prokaryotic ssDNA replicons, such as the bacteriophage ϕX174, it has been suggested that geminiviruses have a prokaryotic origin (Koonin and Ilyina, 1993; Rigden *et al.*, 1996). Observations lending support to this hypothesis are the detection of replicative form (RF) AbMV DNA in chloroplasts, implying that this virus can replicate in plastids (Groning *et al.*, 1987, 1990), and the fact that ToLCV, TYLCV and ACMV can replicate efficiently in *Agrobacterium tumefaciens* and to low levels in *Escherichia coli* (Rigden *et al.*, 1996; Selth *et al.*, 2002).

Since geminivirus multiplication is heavily reliant on the host DNA replication machinery (Rep being the only virus-encoded protein indispensable for the process), this implies that geminivirus genomes have retained the capacity to be replicated by prokaryotic enzymes. However, this convincing argument has been complicated by the recent discovery of a category of eukaryotic DNA transposons (called *Helitrons*), which transpose by RCR in *Arabidopsis thaliana*, *Oryza sativa* (rice) and *Caenorhabditis elegans* genomes (Kapitonov and Jurka, 2001). *Helitrons* encode a replication initiator protein similar to the Rep from other RC replicons, as well as their own helicase and single-strand binding protein (SSB; in plants only). This structure is more like that of geminivirus Reps, which also have a putative helicase domain, while the Reps of prokaryotic RC replicons are usually assisted by host DNA helicases and SSBs. This finding suggests that geminiviruses might have evolved from plant RC transposons rather than from prokaryotic RC replicons (Kapitonov and Jurka, 2001).

Adding to the debate on the origin of geminiviruses, and of eukaryotic ssDNA viruses in general, is the discovery of a plasmid, obtained from a phytoplasma, that encodes a Rep protein whose N-

terminus has similarities to the Rep of prokaryotic RC replicons, while its C terminus is similar to the helicase domain of the Rep of eukaryotic viruses, especially circoviruses (Oshima *et al.*, 2001). The authors speculate that this phytoplasma plasmid may either be an ancestor of eukaryotic ssDNA viruses, or intriguingly may have arisen out of a recombination event between a prokaryotic plasmid and a eukaryotic virus.

Even more surprising information on the evolutionary history of viruses came from comparative analysis of the 3D NMR structure of the catalytic domain of TYLCV Rep with other viral proteins, which revealed a conserved architecture for a number of functionally diverse proteins (Campos-Olivas *et al.*, 2002). These include the RNA binding domain from U1A and other RNA-binding proteins, and the DNA-binding domains from SV40 T-ag and E1 and E2 from papillomaviruses. This structural conservation suggests there is an evolutionary relationship between primordial ssRNA-binding proteins, RCR initiator proteins (prokaryotic and eukaryotic), and mammalian tumour virus proteins such as SV40 T-ag. Accordingly, the development of ancient RCR elements to the more sophisticated DNA tumour viruses mirrors the evolution of their hosts (Campos-Olivas *et al.*, 2002). No doubt further investigations will reveal more interesting links in the evolutionary histories of viruses and plasmids.

1.2.2 The *Maize Streak Virus* Genome

The MSV genome is deceptively simple: at ~2.7 kb and with only four genes, it is one of the smallest virus genomes known. However, for this very reason, the regulation and interactions of MSV genes are extremely complex. To perform the tasks required by the virus to establish an infection, MSV genes and their products must have several distinct activities rather than each gene encoding a product with a single function. They must also be capable of interaction with each other, with viral DNA and with host factors, and have a role in regulating the virus life cycle. As will be seen in the following description of the MSV genome, MSV *Rep* is a perfect example of multifunctionality, playing a pivotal role in the virus life cycle.

1.2.2.1 The long intergenic region

The MSV LIR contains bidirectional C and V sense promoters and associated transcriptional regulatory elements, sites for the binding of Rep and plant nuclear factors for gene expression and replication, and the plus strand (V sense) origin of replication. The latter is found in the highly conserved nonanucleotide sequence TAATATTAC, situated in the loop of a stem-loop structure.

The plus-strand nonanucleotide sequence, present in all geminivirus IRs, is very similar to that found in the origins of nanoviruses (Bell *et al.*, 2002), circoviruses, and prokaryotic RC systems (Baas, 1987). The stem-loop structure in all geminivirus IRs is essential for replication; however, the sequence of the stem is not so important as long as the structure is maintained (Orozco and Hanley-Bowdoin, 1996; Willment, 1999). The sequence does however contribute to the efficiency of replication, as stem variants replicate less efficiently than wild type structures (Orozco and Hanley-Bowdoin, 1996; Schnippenkoetter *et al.*, 2001). The loop can also tolerate some mutations (Schneider *et al.*, 1992; Stanley, 1995); however the loop's nonanucleotide sequence is less flexible: insertions and deletions are not tolerated, although in MSV a mutation of TAATATTAC to TAATACTAC resulted in only a slightly less fit virus (Schneider *et al.*, 1992). Apart from the one notable exception of WDV, the stem-loop structure appears to be required for both initiation and termination of replication. Surprisingly, a WDV mutant with a deletion of the entire stem-loop sequence has been shown to initiate replication. There was, however, a defect in termination of replication, resulting in high molecular weight concatemeric forms of viral DNA (Kammann *et al.*, 1991). Heyraud *et al.* (1993b) subsequently found that the mutant virus was able to initiate replication from a second site (TACCC) resembling the nicking site in the stem-loop. However, this second initiation site is apparently unique to WDV, and it is likely that the stem-loop is required for replication initiation and termination by all other geminiviruses.

In order to understand the features of the mastrevirus LIR, it is helpful to first summarise the more-intensively studied IR of begomoviruses, also known in bipartite begomoviruses as the common region, or CR.

The begomovirus plus strand origin was first mapped by Lazarowitz *et al.* in 1992, and then further defined by Orozco *et al.* (1998), to an 89-bp sequence in the CR. This fragment, which is conserved between the A and B components of begomoviruses, includes the stem-loop sequence and an adjacent 60 nt located on the left side (5') of the CR (Fig. 1.2). Around the same time, Fontes *et al.* (1992 and 1994a) showed that TGMV Rep binds specifically to a 13-bp directly repeated motif (5'-GGTAGTAAGGTAG) that is essential for replication. This site, located 34 bp upstream of the hairpin and between the TATA box and transcription start site of the *Rep* promoter, also mediates transcriptional repression of the Rep gene (Eagle *et al.*, 1994), possibly through interference with assembly or activity of the transcription pre-initiation complex. In this way, TGMV Rep negatively regulates its own expression. Related virus-specific Rep-binding

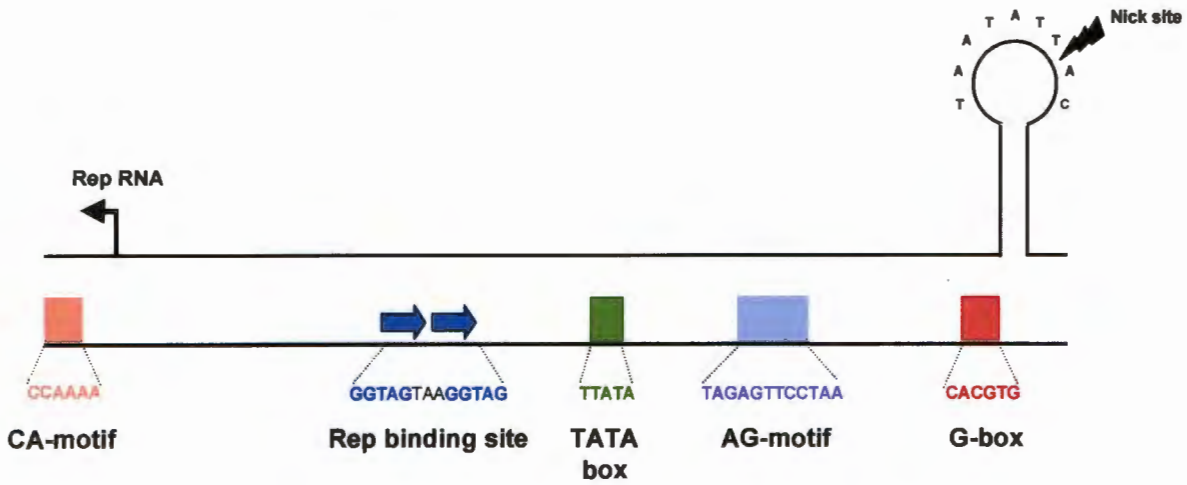

FIGURE 1.2 A schematic representation of the six known functional elements of the *Begomovirus* plus strand origin of replication. These are: (1) a conserved nonanucleotide motif in the loop of a stem-loop structure, which contains the replication initiation site; (2) a G box 5' of the stem which binds host transcription factors; (3) an AG motif between the G-box and (4) the Rep TATA box, which may bind host factors to facilitate initiation of replication; (5) a high affinity Rep-binding site between the *Rep* TATA box and transcription start site, which confers replication specificity; and (6) a CA motif upstream of the Rep binding site, which contributes to efficient replication possibly by binding host factors. The two repeats within the Rep-binding site have different functions, the 3' repeat being essential while the 5' repeat enhances Rep binding (Fontes *et al.*, 1994a).

sites are found in the IRs of all other begomoviruses (Arguello-Astorga *et al.*, 1994a; Fontes *et al.*, 1994b) and curtoviruses (Choi and Stenger, 1996), although the distance between the Rep-binding site and the stem-loop in different viruses is varies from 23 to 82 bp (Palmer and Rybicki, 1998). These Rep binding sites act as origin recognition elements to allow virus-specific replication.

Further analysis of the IR sequences from different begomoviruses (Arguello-Astorga *et al.*, 1994a and 1994b) revealed a series of repeated sequence elements of 8 - 12 nt, called iterons, which have sequence homology to the Rep-binding site identified by Fontes *et al.* (1994a, b). The nucleotide sequence of the iterons is virus specific, but the organization (number, orientation and spacing) is conserved within the CR of dicot-infecting begomoviruses. Based on these observations, Arguello-Astorga *et al.* (1994a) proposed that Rep binds specifically to these sites, and that they have a role in viral replication and/ or transcription.

The distance of the Rep-binding sites from the replication initiation site in the hairpin loop raises the question of how Rep catalyzes cleavage in the loop while bound to its distantly located binding sites. In their computer analysis of the begomovirus IR, Arguello-Astorga *et al.* (1994a)

identified a G-box-like site adjacent to the 5' end of the stem-loop (Fig. 1.2). G-boxes are known to be transcriptional *cis*-regulatory element in plants. Taking into account that the iterons are closely associated with the Rep TATA box, the authors proposed a model by which initiation of replication can occur: A host transcription factor bound at the G-box adjacent to the stem-loop interacts with the TATA-binding protein (TBP), resulting in sequence looping, bringing the Rep complex bound at the TATA-proximal iterons in contact with the cleavage site of the loop to initiate RCR. This process may be assisted by the replication enhancer protein, REn, which has been shown to interact nonspecifically with Rep (Settlage *et al.*, 1996) and possibly to increase the affinity of Rep for the origin (Fontes *et al.*, 1994a; Gladfelter *et al.*, 1997). One proposal is that REn bound to the stem-loop could recruit Rep to the nonanucleotide cleavage site (Hanley-Bowdoin *et al.*, 1996). However, there is no evidence for specific REn interaction with the stem-loop, and as yet there is only speculation on the mechanism by which the protein enhances replication of begomoviruses and curtoviruses.

In addition to the G-box, two other key elements have been identified in the begomoviral IR (Fig. 1.2). One element, the AG motif, is between the G-box and TATA box, and is essential for origin function (Orozco *et al.*, 1998). The second element, a CA motif, is located outside of the minimal origin immediately upstream of the Rep binding site, and may act as an efficiency element (Orozco *et al.*, 1998).

The six known functional elements in the begomovirus IR contributing to origin function are illustrated in Fig. 1.2. The fact that these elements are closely spaced, and that changes in spacing affects origin activity (Orozco *et al.*, 1998), suggests that they interact with each other during initiation of replication, most likely through the proteins that bind to them (Hanley-Bowdoin *et al.*, 1999).

The IR of mastreviruses differs in several ways from that of begomoviruses. Whereas the stem sequence of the IR stem-loop is highly conserved between different begomoviruses, there is low sequence homology between the LIR stem sequences of different mastreviruses (Rybicki, 1994; Padidam *et al.*, 1995). This suggested to Arguello-Astorga *et al.* (1994a) that the mastrevirus Rep binding sites could reside in the stem; and indeed, sequence analysis of eight mastrevirus LIRs showed that a GC-rich sequence in the stem was iterated elsewhere in the LIR (Fig 1.3 [WDV] and 1.4A [MSV]). The positions of these iterons (one between the TATA box of the Rep gene and its transcription start point, and two on either side of the stem) were conserved in all the

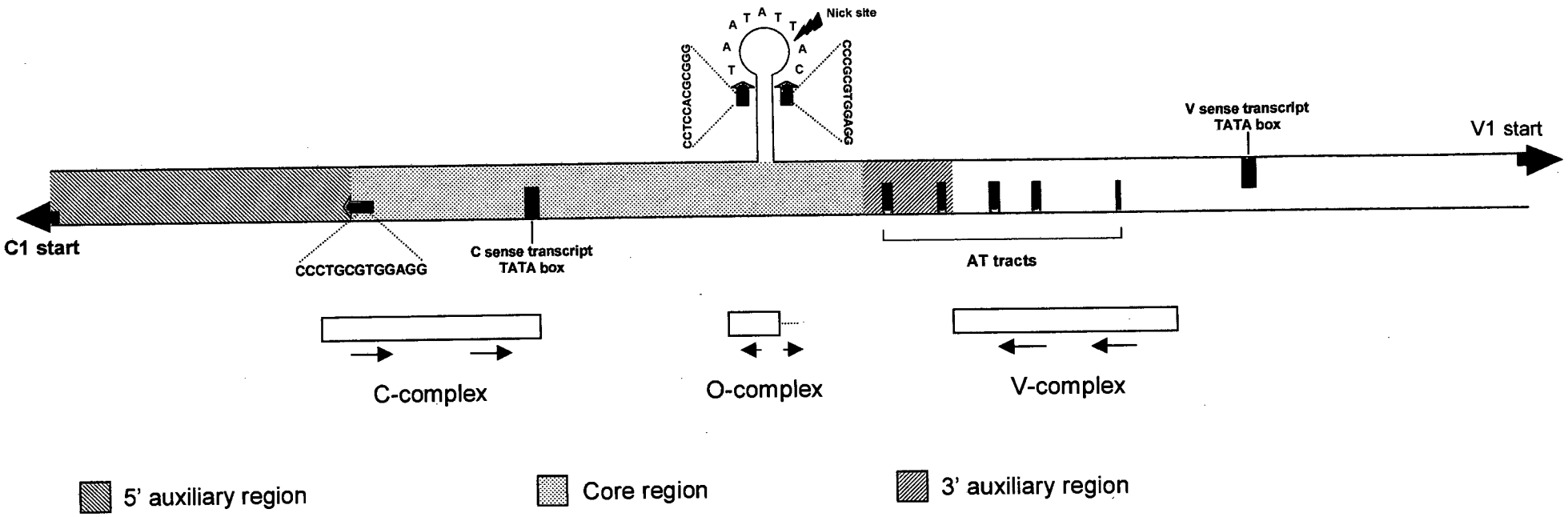

FIGURE 1.3 The main features of the *Wheat dwarf virus* (WDV) long intergenic region (LIR). These include (1) a stem-loop structure with an invariant nonanucleotide sequence where plus strand replication is initiated by an endonucleolytic nick introduced by Rep; (2) the TATA boxes of the two promoter elements controlling C and V sense transcription, 5' and 3' of the initiation site respectively; (3) the start of the C1 (5') and V1 (3') ORFs; (4) iterons represented by blue arrows, with the iterated sequences indicated in the V sense; (5) *cis* elements which include 5' and 3' auxiliary sequences (diagonal stripes) flanking a core sequence element (blue dots) essential for origin function. The core sequence includes the stem-loop structure and invariant nonanucleotide sequence. Also shown are the Rep binding sites defined by mutational, DNase I and footprinting studies. The C and V complexes have high affinity, and the O complex low affinity, footprints. The black (C and V complex) and grey (O complex) arrows indicate G+T-rich direct repeats that may be Rep recognition sequences. Because of their positions within the LIR, it is thought that WDV C and V complexes may be involved in regulation of C and V sense transcription respectively, while the role of Rep bound at the O complex, which is capable of introducing an endonucleolytic nick in the invariant nonanucleotide sequence, may be to initiate plus strand replication

mastreviruses the authors examined. By analogy with the iterated sequences observed in begomoviruses and curtoviruses, the authors proposed that the *Mastrevirus* iterons are also Rep recognition sequences. However, despite some results by Castellano *et al.* (1999; see later in this section) supporting this hypothesis, as yet a direct, specific interaction between iterons and Rep has not been experimentally demonstrated. There is also some evidence that, contrary to the prediction of Arguello-Astorga *et al.* (1994a), in MSV the stem-loop is not a specificity determinant, although a wild type stem sequence does enhance replication (Willment, 1999), a situation similar to that in begomoviruses.

The best-characterised mastrevirus LIR is that of WDV (Suarez-Lopez *et al.*, 1995; Suarez-Lopez and Gutierrez, 1997; Sanz-Burgos and Gutierrez, 1998; Castellano *et al.*, 1999, Missich *et al.*, 2000), the features of which are summarized below and illustrated in Fig. 1.3. Although the WDV LIR differs from that of MSV, some parallels may be drawn in an attempt to fully describe the MSV LIR.

A *cis* element unique to the mastrevirus LIR is a potential static DNA curvature first discovered in WDV (Suarez *et al.* 1995), produced by an 80-bp cluster of "A-T tracts" and located between the stem-loop and MP start codon (Fig 1.3). Since bent DNA sequences have been shown to be necessary for origin activity in a number of systems, the authors hypothesized that the bending locus could be a regulatory element of WDV replication. However, it was later shown that the DNA-bending locus has only a minimal impact on the replication of WDV (Suarez-Lopez and Gutierrez, 1997). It is more likely that bending of this region is involved in regulation of transcription of V sense genes (Castellano *et al.*, 1999).

The minimal origin of WDV replication (a ~200 bp core) was defined by Sanz-Burgos and Gutierrez (1998) as spanning a region ~170 and 28 bp upstream and downstream, respectively, from the initiation site in the stem-loop. This minimal *cis*-acting element, which includes the iterons identified by Arguello-Astorga *et al.* (1994b), is flanked by two auxiliary elements (5' aux and 3' aux) that enhance WDV replication. While the features of the 5' aux region are not known, the 3' aux region contains part of the AT-rich sequence conferring a static DNA curvature of the LIR (Fig. 1.3). Interestingly, visualization of WDV Rep/DNA complexes in the minimal *cis*-acting region by electron microscopy revealed a high affinity Rep-binding site located upstream of the stem-loop, between the C sense TATA box and transcription start site. This location is very similar to the high affinity Rep-binding site of TGMV identified by Fontes *et al.* (1994b), and

that of BCTV (Choi and Stenger, 1996), suggesting that the general configuration of plus strand origins of all geminivirus genera is similar (Hanley-Bowdoin *et al.*, 1999). Moreover, the binding of Rep in the proximity of the TATA box of the C sense promoter suggests that, like *Begomovirus* Reps, the Reps of mastreviruses may regulate their own expression (Sanz-Burgos and Gutierrez, 1998).

WDV Rep binding sites within the LIR were mapped further by Castellano *et al.* (1999). These authors used electron microscopic visualization and DNAse I footprinting to identify three Rep-DNA complexes, which they named C, V and O (Fig. 1.3). The C complex lies between the TATA box for C sense transcription and the transcription start site (hence the name "C complex"), which is in agreement with the data of Sanz-Burgos and Gutierrez (1998), and similar to the location of TGMV Rep binding sites. The V complex, so called because it is located 52 bp upstream from the TATA box for V sense transcription, interestingly coincides with some of the A-T tracts conferring a static DNA curvature on this region identified by Suarez-Lopez *et al.*, (1995). Since WDV Rep (and/or RepA) has been implicated in up-regulation of V sense transcription (Hofer *et al.*, 1992; Collin *et al.*, 1996), and DNA curvature is known to play a role in transcriptional regulation, it is likely that the V complex is involved in positive transcriptional regulation of the V sense genes. This would partly explain the unique architecture of the mastrevirus LIR (e.g. bending of the LIR DNA and unique Rep binding sites), since in the other geminivirus genera Rep is not directly involved in V sense transcription. Conversely, the role of the C complex may be to down-regulate the expression of Rep; the location of the complex (encompassing both the C sense TATA box and transcription start site) is consistent with this theory. The C complex is also highly likely to have a role in DNA replication, as observed by Sanz-Burgos and Gutierrez (1998), considering its location on the 5' side of the minimal *cis*-acting core sequence required for WDV replication

Microscopic visualization by Castellano *et al.* (1999) of the C and V complexes revealed large spherical nucleoprotein structures, suggesting that the complexes consist of Rep oligomers. This is consistent with the observation that TGMV Rep oligomerizes *in vitro* to form octomeric complexes (Settlage *et al.*, 1996; Orozco *et al.*, 1997) and that MSV Rep monomers interact with one another in yeast (Horvath *et al.*, 1998).

WDV RepA also forms DNA-protein complexes (RepA C and V complexes) in a location similar to those of the Rep-DNA complexes, but with distinct DNaseI footprints (Missich *et al.*, 2000).

The Rep and RepA C complexes are approximately the same size, but the RepA V complex footprint is 10 nt smaller than that of the Rep V complex. In addition, Rep-RepA hetero-oligomers may be involved in complex formation, which is consistent with their dual and possibly co-operative roles in viral replication and transcriptional activation of the V sense promoter. The effect of Rep/RepA homo- and hetero-oligomerization on replication and transcription is discussed in the section on the C sense genes.

Whereas the C and V complexes are high-affinity complexes, Rep interacts with low affinity with WDV DNA in the region of the stem-loop, to form an O complex that is capable of carrying out the cleavage reaction necessary for the initiation of RCR. This experimental result supports the hypothesis of Arguello-Astorga *et al.* (1994a) that the stem-loop of mastreviruses contains specific Rep binding sites. However, although the footprints of the Rep and RepA C complexes also cover an iteron (proximal to the Rep TATA box, i.e. the Rep proximal iteron), the V complex does not encompass any sequence-predicted iterons. Interestingly, the footprints of both C and V complexes cover a G+T-rich repeated sequence, GTGTGAN$_{22-23}$GTG(G)TC that may be the actual Rep recognition sequence (Castellano *et al.*, 1999). A similar, although non-repeated, sequence occurs in the stem (GTGG(T)GG); the fact that it consists of half of the C- and V-complex consensus sequence may be the reason for the low affinity of the O complex (Castellano *et al.*, 1999).

The fact that analogous G+T-rich Rep recognition sequences have not been found in the MSV LIR makes it difficult to predict if similar MSV LIR-Rep complexes form. However, Willment (1999) mapped replication specificity determinants (RSD), which presumably include specific Rep binding sites, to a region spanning the Rep proximal iteron and the stem loop (Fig 1.4A). Interestingly, the major RSD corresponds to the region in WDV covered by the C complex, while a minor RSD is in the vicinity of the O complex, perhaps providing indirect evidence for the existence of these complexes in MSV. Another similarity between the WDV and MSV LIRs appears to be the minimal sequence required for replication, which Willment (1999) found to include sequences 5' of the stem-loop up to and including the Rep proximal iteron, and a 25 bp region 3' of the stem-loop, which is very similar to the WDV minimal *cis*-acting core sequence required for replication delineated by Sanz-Burgos and Gutierrez (1998). However, the essential MSV sequence 3' of the stem-loop does not include the AT tracts that in WDV confer DNA bending. The region 3' of the stem-loop essential for replication appears to be peculiar to mastreviruses, since in begomoviruses only the elements on the complementary side of the stem-

loop are required for replication.

An element found in the MSV LIR that is conserved throughout all the geminivirus genera, is a GC-box (directly repeated in MSV) at the 5' base of the stem (Fig. 1.4A). This binds maize nuclear factors and comprises part of the V sense gene promoter (Fenoll *et al.*, 1990). While the V sense promoter core spans the start of the CP gene through to the 3' side of the stem-loop, optimal expression of the CP requires a region 5' of the stem-loop (530 nt upstream of the CP start codon) called the upstream activator sequence (UAS; Fenoll *et al.*, 1988, 1990). Within the UAS, the region containing the GC-boxes, called the rightward promoter element (*rpe1*), is required for efficient replication as well as transcription of V sense genes, probably by recruiting nuclear factors to the region (Arguello-Astorga *et al.*, 1994b). Transcription of V sense genes is probably directed by two TATA boxes 26 and 214 nt upstream from the MP start codon (Wright *et al.*, 1997). Two 3' co-terminal bicistronic transcripts are produced (one large and one small) from which both CP and MP can be expressed, although CP is more efficiently expressed from the smaller, more abundant, transcript.

While begomoviruses encode a protein – TrAP - that transactivates the CP promoter (Sunter and Bisaro, 1991; 1992; 1997), in mastreviruses it is thought that Rep and RepA act together to activate the CP promoter (Zhan *et al.*, 1993; Collin *et al.*, 1996; Mazithulela *et al.*, 2000; McGivern, 2002). There is far more known about TrAP than the transactivational properties of mastrevirus Rep/RepA. TrAp is a zinc binding protein with an acidic transcriptional activation domain at its carboxy terminus (Hartitz *et al.*, 1999), which activates the CP promoter in mesophyll cells, but acts to de-repress the CP promoter in phloem tissue, and probably interacts with cellular proteins to recognize its target promoters (Sunter and Bisaro, 1997). There is also evidence for tissue specificity of the MSV V sense promoters (Mazithulela *et al.*, 2000; Gooding *et al.*, 1999) but it is unknown exactly what role Rep and/or RepA have to play in conferring the specificity. In MSV, further control of V sense transcription is conferred by the presence of an intron within the MP gene that prevents the production of MP from spliced transcripts, whereas the CP can be expressed from both spliced and unspliced transcripts. It is thought that the intron may enhance expression of the CP gene by an intron-mediated enhancement mechanism, such as that conferred by introns in cereal transgene expression cassettes (Wright *et al.*, 1997). Thus, the ratio of CP to MP is controlled in a number of ways: first, the size of the V sense transcripts determines which gene is more efficiently expressed (CP being expressed from the more abundant shorter transcript); second, splicing of an intron in the V sense transcripts enhances the

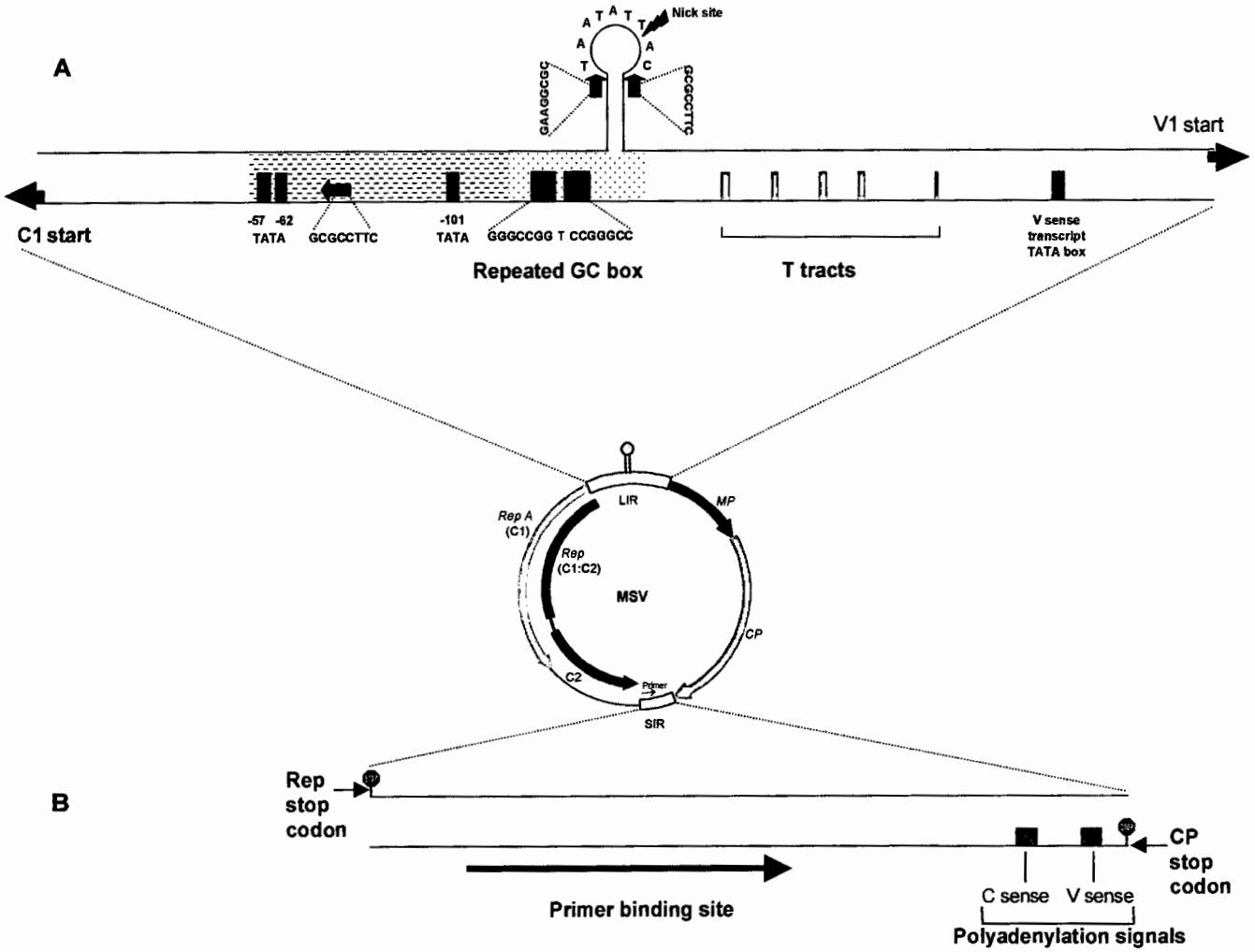

FIGURE 1.4 A schematic representation of the MSV long intergenic region (A) and short intergenic region (B), shown in context with the MSV genome. In (A) the main features of the MSV LIR are shown. These include a stem-loop structure with the loop's nonanucleotide sequence conserved amongst all geminiviruses and other rolling circle systems. The site at which Rep introduces an endonucleolytic nick to initiate plus strand replication is shown. Iterated sequences (iterons) are shown in the V sense, with blue arrows indicating their location in the LIR. Iterons are potentially specific Rep-recognition sequences via which Rep may bind to the LIR. 5' of the stem-loop is a repeated GC-box, which binds host transcription factors, and three TATA boxes from which C sense transcription can potentially be initiated. A series of T tracts 3' of the stem-loop may be involved in DNA bending of this region of the LIR. Also shown 3' is the TATA box for V sense gene transcription. Additional features that may be present in the MSV LIR, but are not shown in (A) include C, V and O complexes that were identified in the *Wheat dwarf virus* (WDV) LIR. These Rep-DNA complexes have not been identified in the MSV LIR. Because of differences in size between the MSV and WDV LIRs, it is difficult to infer analogous positions of Rep binding within the MSV LIR. However, they may correspond to a major replication specificity determinant (major RSD), highlighted in (A) with blue dashes (= C complex?) and a minor RSD, shaded with blue dots (= O complex?) identified by Willment (1999). The minor RSD also includes the stem-loop structure. In (B), the main features of the SIR include polyadenylation signals for V and C sense transcripts, and a primer binding site on the plus strand. A ~80 bp DNA primer-like molecule, encapsidated with the viral genome and annealed to this site, is thought to be involved in initiating negative strand replication. Both the MSV LIR and SIR are essential for viral replication.

production of CP; and third, Rep and RepA appear to transactivate the CP, but not MP, promoter (McGivern, 2002). These areas of control result in CP generally being present at much higher concentrations than MP in infected cells (Mullineaux *et al.*, 1988). The transactivation properties of Rep and RepA are further explored in the section on the C sense genes.

An important difference between the MSV and WDV LIRs is the presence in MSV of three TATA boxes from which C sense transcription can potentially be initiated (Fig 1.4A), located at positions -101, -62 and -57 (relative to the *Rep* ATG): in WDV it can only be initiated from two overlapping TATA boxes at positions -131 and -129. Transcripts initiated at different TATA boxes in MSV could represent one of a few ways in which the relative levels of Rep and RepA are controlled. For example, transcripts initiated from the -62 and -57 TATA boxes are predominantly of 1.2 kb (terminating in the C2), whereas predominantly 1.5 kb transcripts (terminating in the SIR) are initiated from the -101 TATA box (M.I. Boulton, pers. comm). Only RepA can be expressed from the shorter transcript, whereas both Rep and RepA are capable of being expressed from the 1.5 kb transcript. This represents another possible level of control of Rep/RepA expression, whereby splicing of an intron in the larger transcript is necessary to produce Rep, with RepA being expressed from the unspliced 1.5 kb transcript. Thus, modulation of splicing is an obvious way of regulating Rep and RepA expression. Whereas control at the level of splicing can occur in WDV, transcriptional control of Rep/RepA expression may only be possible in MSV. For example, if a Rep C complex does occur in MSV, it is conceivable that binding of Rep oligomers near the -101 TATA box could force C sense transcription from the -57 or -62 TATA boxes, simultaneously enhancing the expression of RepA and inhibiting the expression of Rep. In WDV, however, the C complex encompassing the overlapping –131 and –129 TATA boxes is likely to prevent expression of both Rep and RepA.

Finally, there are differences between the MSV and WDV LIRs in terms of size. The distance between the replication initiation site in the stem-loop and the Rep start codon is larger in WDV than in MSV, as is the distance between the respective Rep proximal iterons and TATA boxes, and the distance of the TATA boxes from the replication initiation sites in each virus. There is also the unique second replication initiation site in the WDV LIR that sets it apart from other mastreviruses. These differences are reflected in the properties of the viral proteins that interact with the WDV and MSV LIRs, as will be seen in the section on the C sense genes, and in the results presented in this thesis. Nevertheless, it is useful to draw analogies from knowledge of the

WDV LIR in order to better understand the interactions of viral and host proteins with the lesser-known MSV LIR. As will be seen throughout this thesis, this is important in determining the multiple functions of the MSV C sense genes.

1.2.2.2 The short intergenic region

The *Mastrevirus* negative (or complementary) strand origin of replication, located in the SIR (Fig 1.4B) and activated by a priming event, is involved at a very early stage of viral DNA replication. Both C sense gene expression and the conversion of plus strand DNA into dsDNA intermediates require activation of the negative strand origin of replication. Essential to this event is a ~ 80bp DNA primer that binds to the SIR plus strand (Fig 1.4B), and is encapsidated with the viral genome (Donson *et al.*, 1984; Hayes *et al.*, 1988; Kammann *et al.*, 1991). In MSV the primer has ribonucleotides linked to its 5' end, suggesting that the DNA is primed from a longer RNA primer molecule (Palmer and Rybicki, 1998). Once virus particles containing plus strand DNA have entered a cell and uncoated, the primer, conveniently already annealed to the plus strand, may initiate synthesis of RF dsDNA, completing the first stage of the viral replication process.

While both the SIR and LIR are required by mastreviruses for efficient genome amplification, in the other geminivirus genera the IR contains the *cis*-acting signals required for both negative and positive strand replication initiation, and therefore constructs containing only the IR support efficient begomoviral replication (Lazarowitz *et al.*, 1992). Although negative strand DNA replication was shown to be RNA-primed in ACMV (Saunders *et al.*, 1992), begomovirus and curtovirus genomes have no virion-associated DNA primer. The origin of the mastrevirus ~80 nt primer-like molecule is unknown, as are the *cis* elements regulating negative sense DNA replication.

Apart from the fact that in addition to its role in replication, the SIR also contains polyadenylation and termination signals of the V and C sense transcripts (Fig 1.4B), little else is known about the shorter of the two mastrevirus IRs. Further elucidation will require more thorough investigation, for example into the minimal sequences required to activate the negative strand origin both within the SIR of mastreviruses and the IR of curtoviruses and begomoviruses, as well as the SIR's potential role in control of mastrevirus gene expression.

1.2.2.3. The complementary sense genes (*Rep* and *RepA*)

One of the earliest steps in the MSV life cycle after passage of the viral genome into the host cell nucleus is the host-directed, DNA-primed synthesis of a complementary (minus) strand, using the virion (plus) strand as the template. This event creates dsDNA, which serves as a template for transcription of viral genes and for RCR. These processes must be tightly controlled, since they cannot occur simultaneously: first the Rep gene must be expressed so that Rep can initiate replication of the genome to a high enough titer to initiate a systemic infection, and only then must the V sense genes be expressed to any significant level. Inappropriate expression of CP and MP would be likely to interfere with viral replication, for example by sequestering ssDNA or by moving the virus genome out of the nucleus. The role of the C sense genes in viral replication, transcription and in regulation of the viral life cycle is covered extensively in this section. First, the biochemical structure (and the functions ascribed to various motifs) of the MSV C sense gene products is discussed. Taken together, this knowledge is used to describe a model of viral replication, for which Rep is the only indispensable viral protein.

There are potentially two transcripts for the MSV C1 and C2 ORFs: these are one of 1.5 kb that results in translation of full-length Rep if spliced, or RepA if unspliced; and one of 1.2 kb that terminates in the C2 and is capable of expressing only RepA . It must be noted that expression of RepA has not been proven *in vivo*; however there is enough evidence of numerous important roles for RepA in the MSV life cycle that in this thesis it is assumed that RepA is an authentic protein.

The MSV C2 ORF, which encodes the carboxy terminus of Rep, is fused with the C1 ORF (encoding the Rep amino terminus) by the splicing of an intron in the C1:C2 transcript. Despite the presence of an ATG start codon in the MSV C2 ORF, it is not considered to encode an autonomous "RepB" protein, partly because there is no detectable transcript for such a protein, and because all other mastreviruses lack a C2 start codon.

As can be seen in Fig. 1.5, Rep and RepA share the same amino terminal 214 amino acids, but differ in their C termini. Although both Rep and RepA appear to be required to activate the promoters of the V sense genes (McGivern, 2002), RepA is unnecessary for viral replication (Schalk *et al.*, 1989; Collin *et al.*, 1996; results in this thesis). However, as will be seen in this section, RepA does appear to play an important role at various stages of the MSV replicative

cycle. These functions, separate and different from those of Rep, are reflected in each protein's unique C terminus.

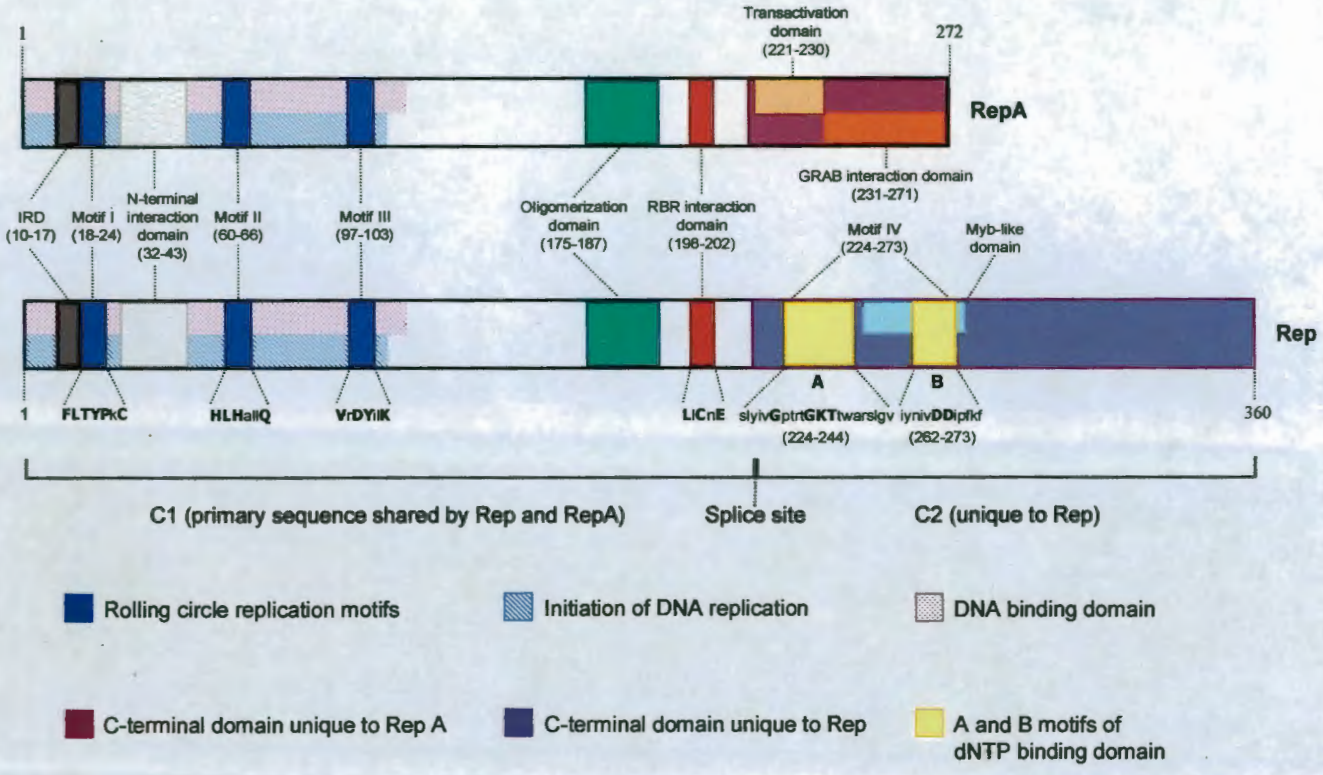

FIGURE 1.5 Predicted functional domains of MSV Rep and RepA proteins (based on Gutierrez, 2000 and Boulton, 2002). The sequences of motifs I, II, III, IV and the RBR protein-interaction domain are shown, with each motif's conserved residues indicated in bold uppercase. The unconserved amino acids (letters in lowercase) are represented by 'x' in the consensus sequences for each motif described in the text. Note that in MSV, the DNA-binding domain and GRAB-binding domain have not been identified, and are shown here in analogous positions to the Rep DNA-binding domain in begomoviruses and curtoviruses, and the GRAB-binding domain of WDV Rep. The N-terminal interaction domain of RepA, although unable to mediate self-association, can interact with the oligomerization domain in the RepA C terminus (Horvath *et al.*, 1998). A RepA fragment containing the N-terminal interaction domain but lacking the oligomerization domain can also interact with full-length Rep through the oligomerization domain of Rep. The biological role of the interaction between the N-terminal and oligomerization domains of both proteins is not clear (Horvath *et al.*, 1998). The details and function/s of each protein domain are discussed in the text

The N terminus of Rep and RepA contains three conserved amino acid sequences (Fig. 1.5) that are related to motifs found in initiator proteins of other RC systems. Although deletion and mutation of each of motifs I, II and III blocks the ability of Rep to cleave and replicate viral DNA, the precise roles of motifs I and II are not known. A tyrosine residue in motif III (VxDYxxK) participates in phosphodiester bond cleavage at the V sense origin of replication (the

loop in the MSV LIR stem-loop structure), and in the covalent linkage of Rep to the 5' terminus exposed by nicking (Laufs *et al.*, 1995b). The nicking-joining activity of geminivirus Reps is the main function of all replication initiator proteins from RC systems. The point of Rep-mediated nicking in the loop is between nucleotides 7 and 8 of the conserved nonanucleotide sequence: TAATATT$^{8\downarrow}$AC (Heyraud-Nitschke *et al.*, 1995; Laufs *et al.*, 1995a; Stanley; 1995). Interestingly, although Rep binds to dsDNA to initiate replication, Laufs *et al.* (1995a) showed that the protein was unable to cleave a dsDNA origin *in vitro*. Although the authors state that there may have been problems in their experimental protocol, this result is corroborated by the finding of Orozco and Hanley-Bowdoin (1996) that formation of a stem-loop structure is required for viral replication. Presumably, transient melting to allow extrusion of a stem-loop structure would produce a ssDNA cleavage substrate in RF DNA.

Motif II (HLHxxxQ) may be involved in metal ion coordination through the histidine residues (Koonin and Ilyina, 1992); the finding that the covalent linkage of Rep to the 5' terminus of the nicked nonanucleotide motif *in vitro* requires Mg^{2+} or Mn^{2+} ions (Laufs, 1995a) supports this hypothesis.

Although no precise function has been ascribed to the conserved FLTYPxC signature of motif I, Arguello-Astorga and Ruiz-Medrano (2001) presented evidence that the region containing motif I may be involved in Rep recognition of iterons in the IR of geminiviruses. The most significant feature of this motif I-associated subdomain, which the authors called an iteron-related domain (IRD; see Fig.1.5), is that its primary structure differs among viruses harboring distinct iterons, while it is generally similar among viruses with identical iterons regardless of their differences in host range, insect vector, geographical origin or genome structure. The authors suggest that, together with motif I, the IRD may form the core of a DNA-binding domain whose secondary structure is apparently conserved in the replication proteins of nanoviruses, circoviruses, microviruses and archaebacterial and eubacterial ssDNA plasmids. Indirect evidence for the IRD being a major component of the specific DNA recognition domain of geminivirus Rep, comes from the fact that the IRD is located within the Rep region where trans-acting replication specificity determinants have been mapped (see references within Arguello-Astorga and Ruiz-Medrano, 2001), and that deletion or mutation of the IRD-Motif I region of TGMV Rep eliminates its specific DNA-binding capability (Orozco *et al.*, 1997; 1998). More recently, the NMR-derived structure of the catalytic domain of TYLCV Rep has added structural evidence for this theory (Campos-Olivas *et al.*, 2002).

A fourth motif ($EGX_4GKTX_{32}DD$), conserved in the C2 of all geminivirus Reps (but absent from RepA; see Fig. 1.5) is a NTP-binding domain with typical A and B motifs that are found in proteins with kinase and helicase activities (Gorbalenya *et al.*, 1989). This motif exhibits ATPase activity that is required for replication, since mutations altering the lysine residue in the P loop of the TYLCV Rep NTP-binding domain impaired ATP hydrolysis *in vitro* and replication *in vivo* (Desbiez *et al.*, 1995). However, neither ATP binding nor hydrolysis is required for nicking and joining of ssDNA at the plus strand origin (Heyraud-Nitschke *et al.*, 1995), or for Rep-mediated transcriptional repression (Eagle *et al.*, 1994). Although a definite function has not been attributed to the NTP-binding domain, it has been speculated that it could enable Rep to act as a helicase that would unwind and displace positive strand DNA from the negative strand template in advance of the replication fork, or unwind and expose the origin to proteins of the replication apparatus (Bisaro, 1996). *In vitro* evidence for a topoisomerase function of Rep was provided by Pant *et al.* (2001). However, because the ATPase activity is DNA-independent (Desbiez *et al.*, 1995), it is still not certain whether its role *in vivo* is to contribute to the putative helicase activity of Rep or another, as yet undetermined ATPase-dependent activity of Rep.

As inferred when discussing motif I, the DNA recognition and binding domain resides in the N-terminus of Rep. While the specific location of this domain has not been mapped in MSV, in begomoviruses and curtoviruses the N-terminal 116 and 89 amino acids respectively are required for Rep to recognize specific DNA-binding sequences in its cognate origin. As mentioned in the section on the *Mastrevirus* LIR, MSV Rep and RepA also bind the origin DNA in a sequence-specific manner, and it is likely that the DNA-binding domain also resides in the N-terminus of both proteins (Fig. 1.5).

In addition to its catalytic and DNA binding activities, mastrevirus Rep is involved in several protein interactions, including homo-oligomerization, binding to RepA, and interaction with host proteins. These interactions and their established or putative roles in the mastrevirus life cycle are described below.

The various activities associated with geminivirus Reps appear to be determined in part by the aggregation state of Rep (Orozco *et al.*, 2000) and in the case of mastreviruses, Rep and/or RepA (Horvath *et al.*, 1998; Missich *et al.*, 2000). Studies have shown that TGMV Rep can form oligomers in solution (Settlage *et al.*, 1996; Orozco *et al.*, 1997), that MSV Rep monomers self-interact in yeast (Horvath *et al.*, 1998), and WDV Rep oligomers consisting of six to eight

monomers have been visualized bound to DNA (Sanz-Burgos and Gutierrez, 1998; Castellano *et al.*, 1999). Although Rep monomers can perform the cleavage and joining reactions *in vitro*, studies with TGMV have suggested that Rep-Rep interaction is a prerequisite for DNA binding (Orozco *et al.*, 1998). Furthermore, in a study of TGMV mutants defective for oligomerization, Orozco *et al.* (2000) discovered a clear correlation between the aggregation state of Rep and its function in replication and transcription. Mutations abolishing or impairing Rep-Rep interaction had the effect of inhibiting or severely impairing TGMV replication, at the same time enhancing Rep-mediated transcriptional repression of the C sense promoter. Interestingly, there are also examples of point mutations in papillomavirus E1 and E2 proteins that differentially affect replication and transcription (Cooper *et al.*, 1999). Replication initiation factors generally function as large protein complexes, whereas transcription factors frequently act as dimers or tetramers (Orozco *et al.*, 2000). Thus, it is conceivable that in geminiviruses different Rep complexes may be required for the two activities. Alternatively, mutations in the Rep oligomerization domain may result in a conformational change that favours repression complexes over replication complexes. For example, a conformational change might make a region that contacts the transcription apparatus more accessible, thus facilitating active repression (Eagle and Hanley-Bowdoin; 1997).

TGMV Rep also interacts with the viral replication accessory factor REn, which enhances viral DNA accumulation. REn can self-interact and oligomerize with Rep in a non virus-specific manner (Settlage *et al.*, 1996). Since the REn protein sequence does not appear to contain any homology to known enzymatic motifs, it is thought that the structure of the Rep/REn complex, rather than a catalytic activity of REn, may be important for replication (Hanley-Bowdoin *et al.*, 1999). Since experiments have indicated that REn increases the affinity of Rep for the origin (Fontes *et al.*, 1994a; Gladfelter *et al.*, 1997), possible functions of REn include (1) REn may direct Rep to its cleavage site in the origin, which in begomoviruses and curtoviruses is located distal from the Rep-binding site (analogous to the effect of papillomavirus E2 on E1 (Mohr *et al.*, 1990); (2) A REn/Rep complex may enhance DNA cleavage and ligation or putative helicase activities of Rep (Hanley-Bowdoin *et al.*, 2000), or (3) REn may play a similar role to transcription factors that enhance replication by stimulating the assembly of the initiation complex on the origin, in which case interaction with Rep would recruit REn to the origin, rather than vice versa. Although none of these possibilities has been proven *in vitro* or *in vivo*, it is very likely that interaction of REn with Rep plays an important role in the mechanism by which REn

enhances replication. Although mastreviruses do not encode a protein with homology to REn, it has been speculated that RepA, which interacts with *Mastrevirus* Rep, could play a similar role.

Yeast two-hybrid studies have identified the domains in MSV necessary for Rep and RepA homo- and hetero- oligomerization (Fig. 1.5; Horvath *et al.*, 1998). Although, as with begomoviruses, oligomerization of mastrevirus Rep is required to assemble a Rep-DNA complex at the origin, in WDV preformed Rep and RepA oligomers bind very poorly to DNA (Missich *et al.*, 2000). This situation is opposite to that reported in TGMV, where it was found that dimerization is required for Rep binding (Orozco and Hanley-Bowdoin, 1998). Interestingly, Missich *et al.* (2000) found that oligomerization of WDV Rep in solution is highly pH dependent within a small, perhaps physiological range. At pH 6.6–7.0, the predominant species is an oligomer (a >6 mer in the case of Rep and an octomer for RepA), while at pH 7.4-7.8 it is a monomer. The fact that preformed oligomers interact poorly with DNA, coupled with the pH-dependent oligomerization property of WDV Rep and RepA, strongly suggests that formation of large oligomers occurs in a stepwise manner. The first stage would be the interaction of a Rep monomer with DNA, mediated by the protein's DNA binding domain. The second stage, requiring the protein's oligomerization domain, would be the sequential addition of Rep monomers, which may lead to the stabilization of the oligomer assembled on the DNA (Missich *et al.*, 2000). Stepwise formation and pH-dependence of oligomeric structures in DNA has been proposed for other proteins, e.g. polyomavirus T-ag (Peng *et al.*, 1998) and SV40 T-ag (Runzler *et al.*, 1987). Furthermore, as has been proposed for geminivirus Reps, oligomerization and assembly of functional higher order complexes is crucial for the many different activities of SV40 T-ag forms (Missich *et al.*, 2000, and references within).

Rep and RepA have been implicated in activation of mastrevirus late gene expression (Hofer *et al.*, 1992; Zhan *et al.*, 1993, Collin *et al.*, 1996; McGivern, 2002), and it is likely that interaction of the two proteins plays an important role in this process. MSV RepA alone was shown to activate transcription of both the *HIS3* and *LacZ* reporter genes in yeast (Horvath *et al.*, 1998), and the activation domain was localized to the C terminus of RepA, which is not present in Rep (Fig. 1.5). However, no Rep/RepA-mediated activation of maize ubiquitin or *Cauliflower mosaic virus* (CaMV) 35S promoters was detected in maize cells (McGivern, 2002), suggesting that RepA does not directly activate plant host gene promoters. The same study found that RepA enhances CP gene expression, but only in the presence of Rep. Since RepA alone has transactivation ability in yeast, Rep may only be required for its replication ability, in order to

amplify the transcription template. However, Rep is likely to be required for additional functions, since in experiments using either Rep mutants unable to support replication, or a reporter construct unable to act as a replicon, expression of GUS in maize suspension cells was increased in the presence of Rep/RepA, but not RepA alone, independent of replication (McGivern, 2002). One possibility is that interaction of Rep with RepA may affect the DNA-binding property of RepA at the V complex. Another is that Rep bound at the V complex could recruit RepA to this site, similar to the proposed function of REn in recruiting Rep to the site of replication initiation. Further proposals include the requirement of a NLS, present in the C terminus of Rep but not in RepA, to target RepA to the nucleus (Boulton, 2002). This would explain why RepA alone can transactivate reporter gene promoters in yeast (Horvath *et al*, 1998), but requires Rep to transactivate the CP promoter in plant cells (McGivern, 2002), since in yeast RepA is directed to the nucleus by the GAL4 nuclear targeting domain.

Rep, however, may simply be required for its own transcriptional activation activity. Two studies found that a region of the Rep C2, which is not present in RepA, can also activate reporter gene transcription in yeast in the absence of RepA (Hofer *et al.*, 1992; Horvath *et al.*, 1998), though Horvath *et al.* (1998) found that a deletion of the Rep C terminal 89 amino acids was required for transcription activation function. Interestingly, the transactivation region coincides with both the NTP-binding motif and a domain that shows homology to the DNA-binding domain of the avian myeloblastosis (*myb*) related class of plant transcription factors (Fig. 1.5; Hofer *et al.*, 1992). Although the transactivation of reporter genes by Rep was detected in yeast, this region may be functionally homologous in maize cells. In this case the activation region could allow Rep to activate either the *CP* promoter, or the promoters of cellular genes required during the replicative cycle.

The fact that full-length Rep is unable to activate reporter gene transcription implies that the activation domain is masked due to protein folding, but is exposed when Rep is truncated. Horvath *et al.* (1998) suggest that interaction between the C terminus of Rep and cellular proteins could expose the activation domain, thus ensuring that transcriptional activation does not occur until Rep has bound to the appropriate host proteins.

It is useful at this point to describe the protein responsible for transcriptional activation of the CP gene in begomoviruses, that is the highly conserved product of the AC2 (or C2) ORF, TrAP. The function of TrAP is not virus-specific, which suggests that either all begomovirus *CP* promoters

contain a common sequence element recognized by TrAP, or that TrAP interacts with cellular proteins common to all begomovirus plant hosts, or a combination of both. The transcriptional activation domain of TrAP, which resides in the acidic C terminus of the protein, also acts as a potent transactivator in mammalian and yeast cells (Hartitz *et al.*, 1999), which suggests that TrAP interacts with factors conserved in yeast, mammalian and plant cells. Optimal interaction of TrAP with viral ssDNA (which occurs in a sequence non-specific manner) requires the binding of zinc, which may facilitate the formation of TrAP-ssDNA complexes (Hartitz *et al.*, 1999). In addition, TrAP is phosphorylated; thus alternative phosphorylation may generate TrAP isoforms that stimulate transcription by different mechanisms. For example, phosphorylation may play a role in the mechanism by which TrAP activates *CP* expression in mesophyll cells, but derepresses the *CP* promoter in vascular tissue (Sunter and Bisaro, 1997). The most likely function of TrAP is to bind and recruit components of the transcription machinery, e.g. general transcription factors (GTFs) and TATA binding-protein-associated factors (TAFs) to the *CP* promoter. It has been shown that GTFs (including TATA binding protein) are contacted by acidic activation domains; thus the TrAP activation domain may stimulate transcription by interacting with similar factors (Hartitz *et al.*, 1999). It would be interesting to determine if the *Mastrevirus* RepA, in combination with Rep, plays a similar role to the begomoviral TrAP in recruiting components of the transcription machinery to the *CP* promoter.

A domain unique to RepA that mediates interaction with cellular proteins is a so-called GRAB (geminivirus RepA binding) protein-binding domain (Xie *et al.*, 1999). By using WDV RepA as a bait in the yeast two-hybrid system, Xie *et al.* (1999) isolated a family of proteins (GRAB), the N-terminus of which exhibits a significant amino acid homology to the NAC (non-apical-meristem, ATAF and CUC2 genes) domain present in a family of plant-specific proteins that are involved in a variety of processes, ranging from lateral root formation to development and senescence. Although this study was done in WDV, the residues required for interaction with GRAB proteins are located in the C terminal domain of RepA (Fig. 1.5), a region that has a significant degree of conservation in all mastrevirus RepA proteins (Xie *et al.*, 1999), except one, that is *Miscanthus streak virus* (MiSV) which lacks the C terminal region (Boulton, 2002). Thus, it is likely that the same interaction occurs in other mastreviruses. The fact that expression of GRAB proteins severely interferes with WDV DNA replication (Xie *et al.*, 1999) points to a role of GRAB proteins in cellular pathways that negatively affect viral replication. By binding to GRAB proteins, RepA may remove the inhibitory block on viral replication. However, a mutant MSV genome expressing a RepA protein that lacks the C terminal 45 amino acids (and therefore

the putative GRAB binding domain) is able to infect maize (Boulton, 2002). Thus, if this domain does exist in MSV, it is dispensable for viral infection of maize.

Further associations of WDV Rep with cellular proteins, identified in a similar manner to that described above, include an interaction with the wheat replication factor C complex (TmRFC-1) (Luque *et al.*, 2002). In eukaryotes RFC is crucial for the recruitment of DNA polymerase δ. During initiation of viral replication, Rep generates a 3'-OH terminus, to which cellular replication factors must be recruited for the assembly of an elongation complex. The data of Luque *et al.* (2002) suggest that WDV Rep, having introduced a nick at the initiation site in the LIR and produced a 3'-OH terminus, stimulates the recruitment of RFC by binding to the RFC large subunit, leading eventually to the recruitment of DNA polymerase δ. This model is further explored in the section on geminivirus replication.

The method employed by Xie *et al.* (1999), using RepA as a bait to bind to host proteins, is a potentially very useful way of identifying cellular proteins involved in processes related to plant growth and the cell cycle, since one of RepA's main functions is thought to be to influence the plant cell regulatory cycle to the benefit of viral replication (Xie *et al.*, 1995). An important domain in *Mastrevirus* Rep and RepA that is the focus of much ongoing research in mastreviruses is the retinoblastoma-related (RBR) protein interaction domain (Fig. 1.5). The retinoblastoma (Rb) tumour suppressor protein, the founding member of the family of so-called "pocket proteins", has long been known to play a pivotal role in the regulation of the human cell cycle. In particular, Rb negatively regulates the cellular G1/S transition of the proliferative cell cycle and is required for proper differentiation of certain cell types. For example in skeletal muscle Rb is required both for expression of late stage differentiation markers and for irreversible exit from the cell cycle (Adams, 2001). The role of Rb is mediated, at least in part, by binding to the E2F family of transcription factors via the Rb "A/B pocket" domain. E2F binding sites have been identified in the promoters of a number of genes involved in DNA replication and progression of the cell cycle. Thus, by binding to E2F transcription factors Rb exerts a block on cell proliferation (de Jager and Murray, 1999). Although *in vitro* Rb can inactivate E2Fs simply by binding and masking the E2F transactivation domain (Ross *et al.*, 1999), Rb also represses transcription by binding to histone deacetylase (HDAC), which removes acetyl groups from the tails of histone octomers (Dahiya *et al.*, 2000). This histone deacetylation activity appears to facilitate condensation of nucleosomes into chromatin, which in turn blocks access of transcription factors, leading to gene repression. Therefore, active repression by the Rb-E2F

complex at the promoters of cell cycle genes is thought to be mediated at least in part by recruitment of HDAC by Rb; an IXCXE site in the C terminus of HDAC seems to be important in mediating association with Rb (Magnaghi-Jaulin *et al.*, 1998).

Rb is regulated by the activity of cyclin/cyclin-dependent kinases (cdks), which are heterodimers of a catalytic cdk and a regulatory cyclin subunit. D-type cyclins, whose expression is strongly induced by growth regulatory signals, bind to and activate their specific cdk partners (particularly cdk4). These kinase complexes interact with the Rb A/B pocket via a conserved N-terminal Rb-binding motif consisting of the amino acids LxCxE in the D-type cyclins. As a result of this interaction, D-cyclin kinases direct the phosphorylation of Rb on multiple cdk phosphorylation sites. Phosphorylated Rb is inactive, since it can no longer bind to E2Fs, and E2F-regulated genes are thereby released from the transcriptional silencing induced by Rb, allowing cell cycle progression into S phase to occur (Weinberg, 1995). In turn, control of the kinase complexes is exerted by cdk inhibitory proteins, which inhibit the activity of the D-cyclin kinases in the quiescent or differentiated state (de Jager and Murray, 1999). Recent evidence suggests that this already complicated sequence of events is even more complex, in that Rb is a multifunctional protein which can inhibit transcription through various mechanisms, and that these functions are progressively and cooperatively inactivated by multiple cyclin/cdk complexes during G1 and S phase (Adams, 2001). However, for the purposes of this discussion, the simpler model will suffice.

The *RB* gene is frequently implicated in tumour formation; indeed it was originally identified because individuals who inherit a mutant copy of the gene are predisposed to develop childhood retinoblastoma (Adams, 2001). The *RB* gene is also associated with a number of other cancers, due to mutation of *RB* or abrogation of its function. As an example of the latter, DNA tumour viruses (such as SV40, adenovirus and oncogenic human papillomavirus (HPV) subtypes) produce oncoproteins that bind to the Rb A/B pocket, displacing E2Fs which then allows E2F-directed transcription of the genes involved in DNA replication and which are necessary for efficient viral replication (Weinberg, 1995; de Jager and Murray, 1999).

Until a few years ago, the powerful role of Rb in mammalian cell cycle control, cellular differentiation and development, together with its involvement in cellular transformation and cancer, were thought to be unique to the vertebrates (de Jager and Murray, 1999). That was until the surprising discovery of homologues of the human Rb protein in maize (Grafi *et al.*, 1996; Xie

et al., 1996; Ach *et al.*, 1997) and subsequently other plants, including tobacco, *Chenopodium rubrum* and *Arabidopsis* (de Jager and Murray, 1999). Analysis of the predicted amino acid sequence of the maize protein, ZmRb, revealed a striking conservation of the domain organization of the plant protein with human Rb and the Rb-related proteins, p107 and p130, particularly the A and B domains of the pocket region, which show 50-65% similarity (Gutierrez, 1998). As described above, this region is necessary and sufficient for Rb to bind to and negatively regulate E2Fs, to bind to and be inactivated by D cyclin-cdk complexes, and to bind to and be sequestered by viral oncoproteins. The conservation of ZmRb with mammalian Rb proteins over the A/B pocket raised the possibility that function may also be conserved at the molecular level (de Jager and Murray, 1999).

Even before the discovery of retinoblastoma-related protein (RBR) in plants, evidence was already accumulating for the existence of an homologous Rb pathway in the regulation of the plant cell cycle. First, a plant D-type cyclin (CycD) homologue was isolated from *Arabidopsis* (Soni *et al.*, 1995). Although the *Arabidopsis* CycDs share only a low level of sequence identity with mammalian D-type cyclins, of particular interest was the presence of the Rb-binding LxCxE motif at the N-terminus of the plant CycDs, which raised the possibility that their target could be a plant Rb homologue. The subsequent isolation of CycD genes from a wide range of plant species, showing that the LxCxE motif is an invariant feature (de Jager and Murray, 1999), supported this view.

The second line of evidence came from geminiviruses, when WDV Rep and RepA were found to contain an Rb-binding LxCxE motif, mutation of which reduced viral replication efficiency (Xie *et al.*, 1995). These findings suggested that Rep and/ or RepA may be required to bind a plant Rb homologue, thereby sequestering it in an inactive form to remove a negative block on viral replication, analogous to the strategy used by animal oncoviruses such as SV40, adenoviruses and HPV.

Although ZmRb was able to bind (albeit weakly) to both human E2F-1 and the Drosophila E2F homologue dE2F, and to negatively regulate E2F-dependent transcription (Huntley *et al.*, 1998), the identification of plant E2F-like proteins proved elusive. Finally, in 1999 Ramirez-Parra *et al.* isolated from wheat a cDNA that encoded an E2F homologue, providing further evidence that the components of the Rb pathway are present in plants. In addition, ZmRb was found to interact much more strongly with this plant E2F than with human E2F-1 (Inze *et al.*, 1999), and putative

E2F binding sites were identified in some plant promoters that show S-phase specific expression (Inze *et al.*, 1999).

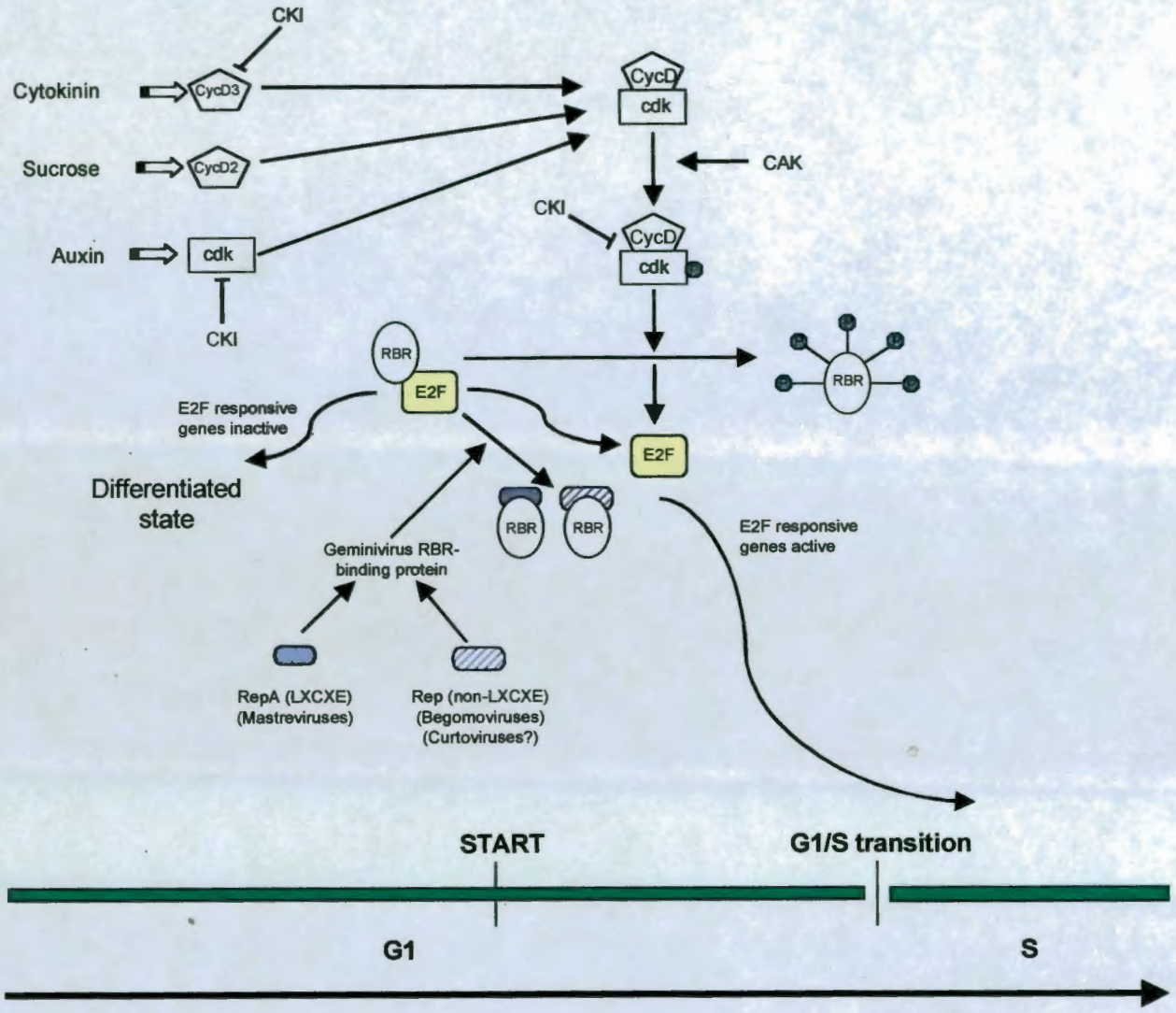

FIGURE 1.6 Model for the control of the G1-to-S phase transition in plant cells, and for the proposed interference of geminivirus Rep proteins with the retinoblastoma-related (RBR) pathway (adapted from the diagrams of Meijer and Murray, 2000; and Gutierrez, 2000). The key components of the pathway are shown. These include CycD cyclins, which are plant homologues of the mammalian D-type cyclins; cyclin-dependent kinases (cdk); cdk inhibitory proteins (CKI); E2F transcription factors that activate the expression of S-phase-specific genes; and retinoblastoma-related (RBR) protein, the plant homologue of mammalian retinoblastoma (Rb) protein. RBR is regulated by the activity of CycD/cdk complexes, whose expression is strongly induced by growth regulatory signals (hormones/ nutrients). The CycD/cdk complex is activated by specific phosphorylation of the cdk component by cdk-activating kinase (CAK), and cdk in turn directs the phosphorylation of RBR. This results in the dissociation of E2F from RBR, allowing expression of S-phase genes and cell cycle progression. The normal RBR phosphorylation pathway may be bypassed in geminivirus-infected cells by the action of RepA (mastreviruses) and Rep (begomoviruses and potentially curtoviruses). By binding to RBR, these proteins may drive cells into S phase (or enable the activation of specific S-phase factors) by promoting release of E2F from RBR.

Further results pointing to a functional conservation of plant and mammalian Rb include the kinase activity displayed against ZmRb by human D cyclin-cdk, that requires an intact ZmRb A/B pocket (suggesting a specific interaction with the A/B pocket is required to exert their kinase function); the binding of SV40 LTA and HPV E7 proteins to ZmRb (de Jager and Murray, 1999); and the binding of *Arabidopsis* CycD cyclins to ZmRb via their LXCXE motif (Huntley *et al.*, 1998). Recently, direct evidence that ZmRb is a cell cycle regulator was provided by Gordon-Kamm *et al.* (2002), who demonstrated that expression of ZmRb inhibits cell division in tobacco cell cultures.

Taken together these results suggest that, although specific mechanistic details may differ (Mironov *et al.*, 1999; Ramirez-Parra *et al.*, 1999), the major components of the mammalian and plant Rb pathways are functionally equivalent. In the plant RBR pathway model (illustrated in Fig. 1.6), growth stimulatory mechanisms such as sucrose and cytokinins induce the expression of CycD cyclins. These form active kinase complexes targeting RBR for inactivation and dissociation from the promoter-bound E2Fs, allowing expression of S phase genes and culminating in DNA replication and cell cycle progression (de Jager and Murray, 1999). The RBR pathway may also be important in plant differentiation and development. For example, in developing maize leaves, which show a gradient of cell proliferation from actively dividing cells at the leaf base to differentiated cells nearer the tip, the ZmRb protein is abundant in the differentiating cells at the leaf tip and almost undetectable in the proliferating cells of the leaf base (Huntley *et al.*, 1998).

The excitement associated with the discovery of an Rb pathway conserved in mammals and plants is understandable, as it may help our understanding of plant and animal evolution. Given that the components of the Rb pathway are not present in yeast or fungi, which use unrelated proteins to control the same processes, a common pathway relating cell division and differentiation may have arisen only once in the evolution of all higher eukaryotes. This suggests that the invention of the Rb pathway may have been the factor that allowed multicellular organisms to develop complex body plans (de Jager and Murray, 1999).

The discovery of the LxCxE RBR protein interaction motif and other cellular protein-binding domains (such as the GRAB-binding domain) in WDV RepA has prompted some intense research, with the aim of not only elucidating the components and mechanisms of the viral life cycle, but also of providing insights into DNA replication, cell cycle and growth control in plants.

Now that a background has been given on the Rb protein function and the discovery of its homologue in plants, the relationship between the RBR protein, geminivirus Reps and associated effects on the plant cell cycle in geminivirus-infected plants can be described.

Geminivirus replication depends, apart from Rep, entirely on host cell proteins. These cellular replication proteins are generally absent or not functional in differentiated cells. Geminivirus replication is excluded from the meristem, which contains actively proliferating cells that naturally express the replication factors required for geminivirus replication. Thus, even before the relationship between the plant RBR protein and the WDV Rep LXCXE motif was known, it was suggested that geminiviruses might induce a cellular state permissive for viral replication. There were two main observations supporting this idea. The first was that dsDNA replication intermediates are significantly more abundant in S-phase nuclei than in nuclei from other phases of the cell cycle (Acotto *et al.*, 1993), suggesting that cellular DNA replication (the S phase of mitosis rather than actual cell division) is important for viral replication. The second observation was that proliferating cell nuclear antigen (PCNA), which is virtually undetectable in differentiated cells, accumulates to high levels in fully differentiated cells expressing TGMV Rep (Nagar *et al.*, 1995). PCNA, whose expression is regulated by E2F, has an essential role in DNA replication, functioning as a homotrimeric clamp at the origin, which facilitates the recruitment and processivity of DNA polymerase δ (Sever-Chroneos *et al.*, 2001).

However, in some geminiviruses a correlation was not found between viral and host DNA replication. By comparison of the distribution of replicative forms of MSV DNA with the expression of the S-phase-specific host gene, H2b, Lucy *et al.* (1996) showed that host DNA synthesis is not a prerequisite for MSV replication; this lack of correlation was most notable in the developmentally mature leaf laminal tissues where MSV replication could be detected. This suggests that although different geminiviruses may interfere with control of cell proliferation, more than one mechanism or strategy may have evolved. Although it appears that certain geminiviruses such as BCTV do have the capacity to initiate host cell division (causing tumorigenic growths in infected plants; Latham *et al.*, 1997), the data of Lucy *et al.* (1996) suggest that one or more factors associated with dividing cells, rather than cell division itself, may be required for MSV replication.

Given that geminiviruses may induce a cellular state that is permissive for viral DNA replication, either by inducing cells to enter the S phase or by activating some S phase function/s, the

discovery of RBR in plants and the LxCxE RBR-interaction motif in WDV RepA provided a possible mechanism by which this cell-cycle interference could occur (Fig. 1.6). The LxCxE motif is present in the RepA of most mastreviruses, including dicot-infecting members such as BeYDV (Liu *et al.*, 1999a). Although the mastrevirus Rep shares the LxCxE motif with RepA (Fig.1.5), Rep does not interact with RBR (Horvath *et al.*, 1998; Liu *et al.*, 1999a; results in this thesis), most likely due to steric hindrance induced by the C terminal domain of Rep (Gutierrez, 2000). This raises the possibility that RepA may be required in the viral life cycle for its ability to provide a cellular environment competent for viral replication through interaction with RBR and other cellular factors, such as the GRAB proteins mentioned earlier. There is convincing evidence supporting this view. First, both BeYDV (Liu *et al.*, 1998), and MSV (Boulton, 2002) mutants unable to express RepA are also unable infect plants. Second, expression of WDV RepA stimulates cell division in tobacco cell cultures (Gordon-Kamm *et al.*, 2002). The stimulatory effect is observed not only in mitotically active cells, but RepA also appears to overcome cell arrest in the G0/G1 phase of the cell cycle. Furthermore RepA (but interestingly not Rep) expression stimulates maize embryogenic callus growth and increases transformation efficiency, most likely due to stimulated cell division. Overexpression of ZmRb suppresses RepA-stimulated cell division, suggesting that the effects of RepA on the cell cycle and on transformation are due to the removal of RBR protein-mediated repression of the cell cycle. However, no mastrevirus-induced host cell proliferation has yet been observed, so it is unlikely that in a natural infection these viruses induce total progression from G1 phase through S phase to mitosis.

RepA-RBR interaction may not only have an effect on viral replication. As mentioned previously, RepA has been implicated in the transactivation of the mastrevirus *CP* promoter. This activity could be facilitated by the binding of RepA to RBR protein, thus releasing E2F transcription factors that could then be involved in activating the *CP* promoter. Interestingly, the E1a protein from adenoviruses uses this strategy to promote activation of viral promoters (La Thangue, 1994).

Mutational analysis of the mastrevirus RepA LxCxE motif has shown the importance of the three conserved residues in mediating binding to RBR. In one study, a mutation of C to G had a significant effect on the ability of WDV RepA to bind to Rb, while changing the E residue to K abolished binding (Xie *et al.*, 1995). In a detailed mutational analysis of the motif in BeYDV RepA, Liu *et al.* (1999a) confirmed the importance of all three conserved residues: while mutants containing changes of L to I; C to S; and C to G retained the ability to bind to RBR, they did so

with reduced efficiency. Furthermore, in a result similar to that of Xie *et al.* (1995), mutation of the E residue (in this case E to Q) had the effect of drastically reducing binding efficiency, by 95%. The LXCXE motif also mediates binding of a nanovirus gene component (Clink, for "cell cycle link") to RBR (Aronson *et al.*, 2000), and the importance of the motif in the oncoproteins of animal tumour viruses has already been mentioned. Thus, it was surprising to find that TGMV Rep, which encodes no LXCXE motif, interacts with RBR (Ach *et al.*, 1997) through a different motif (Kong *et al.*, 2000). There are several lines of evidence that TGMV alters the cell-cycle controls of its host plant, *Nicotiana benthamiana*. Indications that TGMV may trigger re-entry of mature cells into the cell cycle include the incorporation of high levels of bromodeoxyuridine into both viral and host DNA in infected cells, suggestive of progression into S phase and DNA replication (Egelkrout *et al.*, 2001), and the fact that a large fraction of TGMV-infected cells contains condensed chromatin, which is characteristic of early mitotic prophase (Bass *et al.*, 2000). The importance of TGMV Rep-RBR interaction was established by Kong *et al.* (2000), who found that mutants with impaired binding to RBR accumulate less viral DNA and cause chlorosis that is confined to the veins, and that the mutations result in altered tissue specificity.

A clue to the mechanism by which binding of TGMV Rep to RBR could alter the cellular environment came from analysis of the *N. benthamiana* PCNA promoter. As mentioned previously, PCNA is an S phase gene that is regulated by E2F. In a study of transgenic plants carrying a mutation of an E2F consensus element in the *N. benthamiana* PCNA promoter, Egelkrout *et al.* (2001) demonstrated that E2F functions as a negative regulatory element to repress PCNA transcription in mature leaves. Thus, E2F can repress as well as activate PCNA promoter activity. The mechanism by which this occurs may be similar to the E2F-mediated control of the tobacco ribonucleotide reductase small subunit gene (RNR2) expression. The RNR2 promoter contains multiple E2F elements, two of which activate transcription and one that represses transcription out of S phase (Egelkrout *et al.*, 2001). It has previously been shown that PCNA expression in infected cells is tightly linked to the capacity of TGMV Rep to interact with RBR (Kong *et al.*, 2000). Taken together, these data suggest that TGMV Rep may disrupt the RBR/E2F complexes that repress PCNA promoter function and in so doing induce PCNA transcription in mature leaves (Egelkrout *et al.*, 2001).

Rep is not the only TGMV protein to interact with RBR protein. Settlage *et al.* (2001) established that the TGMV replication enhancer protein, REn, binds to RBR protein as well as Rep. However, unlike Rep, REn alone cannot induce expression of PCNA (Settlage *et al.*, 2001). The

protein domains that mediate the interactions between Rep, REn and RBR overlap, suggesting that REn serves a dual role in enhancing geminivirus replication and that these two functions are co-ordinated by shared-protein domains. Although Rep oligomerization may be required for REn binding (Settlage *et al.*, 2000) and has been demonstrated to be a prerequisite for RBR binding (Kong *et al.*, 2000), oligomerization of REn is not required for interaction with Rep or RBR. Settlage *et al.* (2001) suggest that Rep and REn may serve different roles in the host induction process. One possibility is that REn regulates Rep/RBR interactions through a partially shared protein interaction domain. Both Rep and RBR are predicted to bind to REn within its first 35 amino acids (Settlage *et al.*, 2001). Given that RBR is greater than 100kDa in size and that Rep probably binds REn as an oligomer, REn may not be able to interact simultaneously with both proteins. The authors propose that Rep/REn, but not Rep/RBR complexes are functional for initiation of geminivirus replication. According to this model, REn modulates the stoichiometry of different Rep complexes, and hence Rep activity, through its binding to both Rep and RBR protein. It would be interesting to determine if RepA, which also binds to both Rep and RBR, has a similar function in mastreviruses.

Apart from RBR protein, geminivirus Rep may also interact with other host proteins involved in cell division and development. For example, interaction of TGMV Rep with a kinase and a kinesin from *Arabidopsis*, both of which are potentially involved in the progression of the cell cycle through the G2 and M phases, possibly prevents transit through the G2 phase, stalling the cell cycle in an S-like phase that results in endoreduplication rather than cell division (Kong and Hanley-Bowdoin, 2002).

Further interactions with plant factors were determined in the nanovirus *Faba bean necrotic yellows virus* (FBNYV) genetic component, Clink. In addition to the LxCxE RBR protein interaction motif, Aronson *et al.* (2000) identified in Clink an F-box that binds to a plant SKP1 homologue, a constituent of the ubiquitin-protein turnover pathway. F-box proteins serve as substrate-specific adapter subunits to recruit various substrates to be ubiquitinated, in preparation for degradation by the 26S proteosome. Therefore, FBNYV Clink may target RBR protein for ubiquitin-mediated degradation, as was described for HPV-16 E7 protein (Boyer *et al.*, 1996). As suggested for TGMV and other geminiviruses, nanoviruses probably achieve viral replication without the completion of mitosis (for example by the process of endoreduplication), since neither group causes uncontrolled host cell proliferation. Thus, targeting the degradation of proteins (including the degradation of Clink itself) in a later phase of the cell cycle could restore a

semblance of normal cell cycle control in the host cells, once a critical amount of viral genome products has accumulated (Aronson *et al.*, 2000).

It would seem that interference with the RBR pathway is an important part of the geminivirus life cycle. However, there is substantial evidence to suggest that the RBR-binding activity of Rep is dispensable for both viral replication in cultured cells and infectivity in host plants. Although Xie *et al.* (1995) found a correlation between the ability of WDV RepA to bind to RBR and the replication efficiency of the virus in wheat suspension cells, no such relationship has been detected in other mastreviruses. Liu *et al.* (1999a) found that all BeYDV mutants, in which RBR binding was impaired by varying degrees up to 95%, were able to replicate in tobacco protoplasts and to systemically infect *N. benthamiana* and bean. In addition, the FBNYV genetic component that binds to RBR, Clink, is not absolutely required for viral replication in *N. benthamiana*, although the LxCxE sequence of Clink does enhance FBNYV replication (Aronson *et al.*, 2000). Further examples of the redundancy of the LxCxE motif in viral genes include those found in potyviruses (a group of plant RNA viruses). Although the potyviral RNA-dependent RNA polymerase (NIb) carries the LxCxE motif, it does not interact with any RBR proteins, and mutation of the highly conserved E residue to K has no effect on viral replication (Oruetxebarria *et al.*, 2002). Although this is not so surprising considering that potyvirus replication is largely independent of the cellular replication activities, the conservation of this motif in a range of plant and animal RNA viruses (a similar motif is also found in the replication proteins of bymoviruses, potexviruses and *Rubella virus*) is suggestive of an alternative function, such as the global folding of the NIb protein necessary for interactions with other proteins during virus replication (Oruetxebarria *et al.*, 2002). Although it may be for a different reason, the possibility exists that in geminiviruses and in particular the mastreviruses, either the LxCxE motif is required for an activity other than RBR binding, or at least one other functional domain may overlap the LxCxE motif, resulting in its conservation in mastrevirus Reps.

Clearly, there is a variety of possible strategies for geminiviruses to interfere with the regulation of the cell cycle and other host pathways, mediated by a number of interactions between viral and host proteins, many possibly as yet undiscovered. Adding to the confusion, the exact function and purpose of the RBR-interaction motifs in geminiviruses, particularly mastreviruses, is yet to be determined. Part of the work in this thesis attempts to provide further insight into the role of the LXCXE motif in MSV Rep.

1.2.3 Replication

Geminiviruses replicate via double stranded circular intermediates, which form minichromosomes within the nuclei of infected cells (Abouzid *et al.*, 1988). The generally accepted model is that geminiviruses replicate using a rolling circle mechanism (RCR). However, recent evidence suggests that geminiviruses might also multiply via recombination-dependent replication (RDR; Jeske *et al.*, 2001). The focus of this section is on the characteristics of RCR, which have been more extensively studied; however it is important to bear in mind that an alternative route of replication by recombination most likely exists.

In the RCR model, amplification of the viral genome occurs in three stages (Fig. 1.7). The first stage is the conversion of the genomic circular ssDNA [(c)ssDNA] into supercoiled covalently closed circular dsDNA [(ccc)dsDNA] intermediates, or replicative form I (RFI). This involves the host-directed, DNA-primed (or RNA-primed in the case of begomoviruses) synthesis of a complementary (minus) strand, as described in the section on the mastrevirus SIR. The second stage is the amplification of the RFI by RCR, which is discussed in detail below. The production and encapsidation of mature genomic (c)ssDNA into viral particles represents the third and final stage of the geminivirus replicative cycle.

The viral RFI serves as a template for both viral transcription (leading early on to the expression of Rep and RepA), and for further DNA replication steps. As mentioned previously, the replication initiation site is contained in the loop of a stem-loop structure within the mastrevirus LIR. DNA sequences around the initiation site comprise the replication origin, which exhibits a modular organization. Upon interaction with specific DNA sequences at the viral origin, Rep introduces a sequence-specific endonucleolytic nick in the V sense genomic strand of the RFI. In the case of WDV, Rep binds with low affinity at the stem of the stem-loop, constituting an O complex, and with high affinity ~140 nucleotides upstream from the initiation site in the vicinity of the C sense promoter, constituting a C complex. Formation of the O complex is sufficient to carry out sequence-specific cleavage at the loop (Castellano *et al.*, 1999). However, it is possible that Rep molecules bound at the C and O complex may interact to form a higher order complex at the origin (Gutierrez, 1999). The initiation reaction involves the nucleophylic attack by the OH group of the conserved tyrosine residue in motif III of Rep, to the phosphodiester bond between the last T and A residues of the invariant nonanucleotide plus strand origin sequence (TAATATT↓AC; Heyraud-Nitschke *et al.*, 1995; Stanley, 1995; Laufs *et al.*, 1995a).

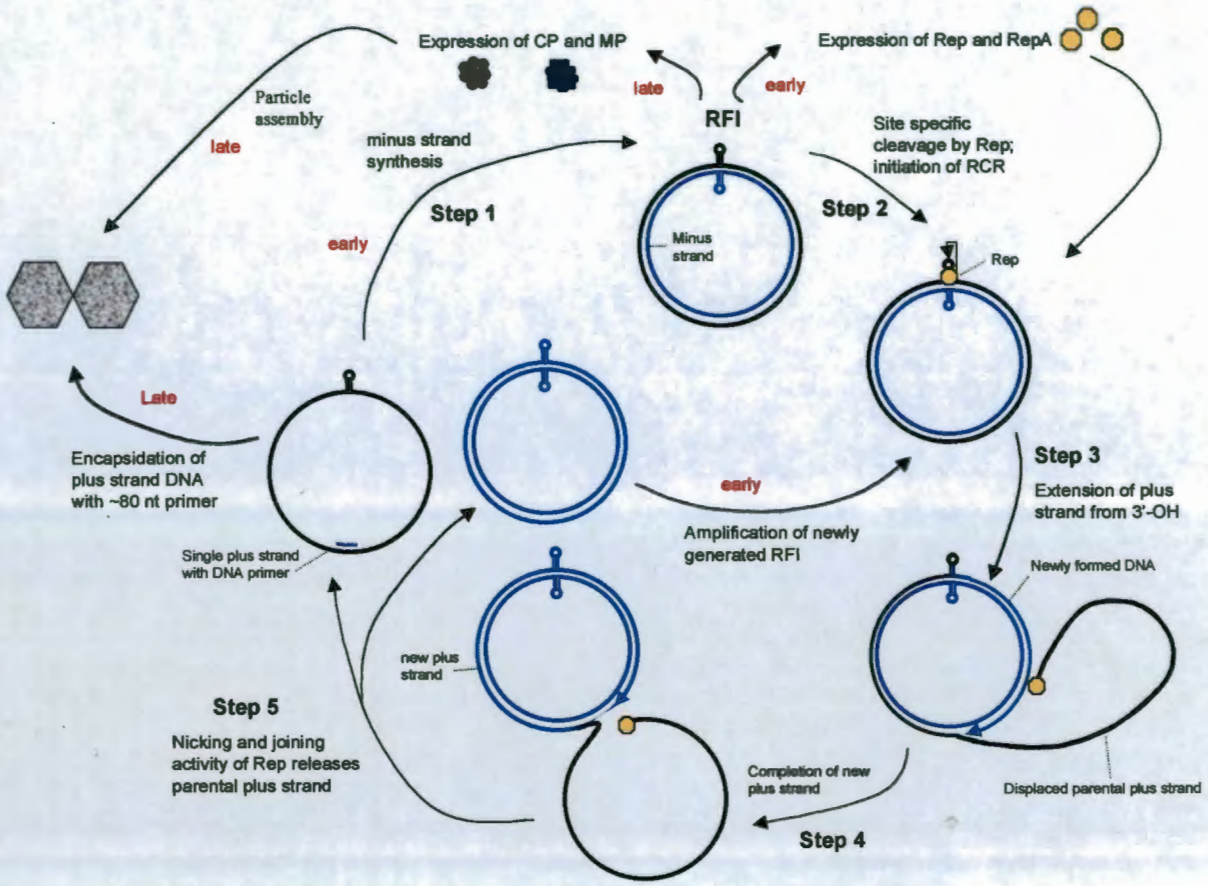

FIGURE 1.7 Summary of the MSV replicative cycle. Early in the cycle genomic circular ssDNA is converted into a dsDNA replicative form intermediate (RFI; step 1). RFI serves as a template for transcription of Rep and RepA (early in the cycle) and CP and MP (late in the cycle). Rep initiates rolling circle replication (RCR; step 2) by binding to the viral origin and introducing a nick in the loop of the plus strand DNA stem-loop structure. The initiation reaction results in the production of a free 3'-OH terminus which is used as a primer for synthesis of a new plus strand (step 3), while Rep remains covalently linked to the phosphorylated 5'-OH end. Once the new plus strand is synthesised by the host replication machinery (step 4), the parental plus strand is displaced from the negative strand template, and termination occurs whereby the nicking-joining activity of Rep simultaneously releases the parental plus strand and liberates a (c)ssDNA molecule (step 5). Early in the cycle the newly liberated plus strand is converted back into RFI to begin the cycle once again, while late in the cycle it is encapsidated and moved out of the cell to establish a systemic infection.

The Rep-mediated initiation reaction results in the production of a free 3'-OH terminus (which is used as a primer for synthesis of a new plus strand), while Rep remains covalently linked to the phosphorylated 5'-OH end (Laufs *et al.*, 1995b). In eukaryotes, the efficient binding of a processive DNA polymerase complex to a 3'-OH primer-terminus depends on the function of the

RFC clamp loader. RFC facilitates loading of the PCNA clamp, which eventually recruits DNA polymerase δ. It is known that in WDV (and most probably other geminiviruses), Rep interacts with the host RFC, thus stimulating its recruitment to the viral origin after initiation. Subsequently, PCNA could be incorporated into the pre-elongation complex, or alternatively, a preformed RFC/PCNA complex could directly be recruited onto the newly formed 3'-OH primer-terminus. Finally, a DNA polymerase would be recruited, concomitantly with ATP hydrolysis and release of RFC, leading to the assembly of an elongation complex that can extend the primer (Luque *et al.*, 2002).

Once the new strand is synthesized, the parental plus strand is displaced from the intact negative strand template, possibly mediated by the putative helicase activity of Rep. The mechanism of termination, whereby the replication cycle is resolved to release a (c)ssDNA molecule, is unknown. By analogy with other rolling circle replicons, geminivirus RCR may be a continuous process, as is the case for phage ΦX174, or a noncontinuous process, as for example in plasmid pC194. In the former case, completion of the new plus strand regenerates the origin of replication, which again is nicked by Rep, this time acting as a terminase to displace the parent plus strand. This Rep becomes covalently linked to the 5'-AC end of the new nonanucleotide, and the 5' phosphoryl group of the displaced strand is then transferred to the newly generated 3'-OH group to liberate a (c)ssDNA molecule. In this model, two active tyrosines are required for switching from initiation to termination. Since Rep contains a single conserved active tyrosine, this mechanism would require two Rep molecules, possibly within the same bound oligomer: in this way an active Rep molecule can always remain attached to the DNA via alternating tyrosines. Alternatively, the mechanism for resolution could be discontinuous, requiring only one tyrosine. As in the continuous model, after nicking Rep becomes linked to the 5' end of the cleaved DNA via a phosphotyrosine linkage. After one round of DNA synthesis, the release of the (c)ssDNA is mediated by a non-tyrosine residue in the same Rep molecule. The newly synthesised origin is then cleaved, and the 5' end that was linked to the Rep active site tyrosine is transferred to the newly created 3'-OH end. Thus, one tyrosine is sufficient to effect nicking and resolution of (c)ssDNA, but the next cycle of replication must be re-initiated by a different Rep molecule, rendering it non-continuous (Palmer and Rybicki, 1998).

Depending on the stage of the replication cycle, the newly released plus strand may either be incorporated into the replication pool to be converted into another RFI (early in the cycle), or will be accumulated as (c)ssDNA, destined for encapsidation or movement to surrounding cells,

thereby removing it from the replication pool (late in the cycle). This aspect is covered in the section on the viral life cycle.

1.2.4 The MSV Life Cycle

An MSV infection is presumably initiated when a leafhopper introduces at least one virus particle (encapsidated ssDNA) into a nucleated plant cell, probably a phloem companion cell. Virus particles delivered into the phloem of leaves are apparently transported within sieve tubes to regions distal from the leafhopper feeding site (Peterschmitt *et al.*, 1992); a sufficient number move out of an inoculated leaf within a few hours to initiate a productive infection (Storey, 1938). It is unknown whether an infection can be established upon entry of a virus particle into any nucleated cell, or whether the cell must be undergoing active cell division and have DNA replication enzymes available. Plant DNA replication and cell division are confined to apical meristems, developing leaves and the cambium of mature plants (Hanley-Bowdoin *et al.*, 1999). Accordingly some geminiviruses, such as AbMV, are restricted to the phloem (Abouzid *et al.*, 1988), possibly replicating in procambial cells. However, other geminiviruses are not confined to vascular tissue, and may modify differentiated cells to induce the synthesis of replication enzymes. For example, MSV DNA can be detected in vascular and mesophyll cells of mature leaves (Lucy *et al.*, 1996), although paradoxically MSV DNA is not detected in meristematic cells.

Once introduced into a permissive cell the first step would be for the virion to uncoat, followed by the movement of the released ssDNA to the nucleus, where viral replication takes place. Nuclear import of viral DNA must then occur in each subsequently infected cell. In MSV, movement of probably partially uncoated ssDNA into the nucleus is facilitated by the CP, which has DNA-binding activity (Liu *et al.*, 1997) and a nuclear localization signal (NLS; Liu *et al.*, 1999b).

Once in the nucleus, the ssDNA viral genome is converted to (ccc)dsDNA (RFI) as described in the previous section. It is assumed that no modification of host cell gene expression is necessary at this stage, since nick repair enzymes and cofactors presumably responsible for synthesizing the negative strand are constitutively expressed (Palmer and Rybicki, 1998).

Once transformed into RFI, the first priority would be to express Rep and RepA. Rep is essential to initiate RCR, and RepA may be required to induce the expression of host enzymes and co-factors required for the completion of RCR. As discussed previously, Rep and/or RepA may directly induce the promoters of certain host genes required for viral replication (since both proteins are transcriptional activators), or they may interfere with cell cycle regulatory systems to indirectly induce the host genes required for virus replication. It is interesting to note that in maize suspension cells the MSV C sense promoter is most active in the early S phase of the cell cycle (before the start of histone H4 transcription), while the CP promoter shows the highest activity in early G2 (Nikovics *et al.*, 2001). The difference in the expression timing from these promoters is consistent with the functions of the MSV gene products. For example, a relatively high level of expression from the C sense promoter by early S phase could result in RepA-mediated manipulation of the host cell cycle by binding to RBR, thereby potentially releasing factors regulating the transition from the G1 to the S phase of the cell cycle. Similarly, Rep is necessary to initiate RCR, which is likely to be accomplished during the S phase. In contrast, CP is required for encapsidation of MSV DNA and systemic infection (Liu *et al.*, 1997) and thus is not needed in the early stages of MSV replication (Nikovics *et al.*, 2001). Interestingly, the C sense promoter is re-activated in the late G2 phase, which may reflect a requirement for Rep or RepA to interfere with progression through G2 phase, thereby locking infected cells into S phase (Nikovics *et al.*, 2001; Nagar *et al.*, 1995).

Apart from their function in the replication of the viral genome, Rep and RepA may have an integral role in the regulation of the entire virus infection cycle. While RepA may be required early in the infection process to prepare the cellular environment for replication, it also potentially plays an important role in activating the CP promoter. Since the CP is only needed later in the infection cycle, this could be another reason for the re-activation of the C sense promoter in the late G2 phase.

Although Rep expression would be expected to rise with the increase in copy number of its gene due to viral replication, Rep transcripts are relatively rare in infected cells (Wright *et al.*, 1997). This implies that Rep expression is tightly controlled. It is known that begomoviral Rep represses its own promoter (Eagle *et al.*, 1994). Since MSV Rep binds near the C sense TATA box (the C complex), it has been suggested that Rep binding in this area interferes with initiation of *Rep* transcription, creating a negative feedback mechanism (Arguello-Astorga *et al.*, 1994a, 1994b). Thus, even though the Rep gene copy number increases exponentially during replication of the

genome, Rep expression is kept fairly constant. However, there is no direct evidence that mastreviral Rep does autoregulate its expression.

RepA has also been implicated in the down-regulation of viral replication (Collin *et al.*, 1996; Liu *et al.*, 1998). Since Rep and RepA have different functions in the mastrevirus life cycle, control of the infection process may be achieved through altering the relative proportions in which Rep and RepA are expressed. This may occur in several ways. Differential splicing is one obvious method of altering the ratios of Rep:RepA. The unspliced transcripts from which RepA is produced comprise approximately 80% of the total C sense transcripts in MSV-infected maize tissues (Wright *et al.*, 1997). Expression of Rep and RepA may also be influenced by the size of the C sense transcript produced: a long transcript (terminating in the SIR) can be translated to produce either Rep or RepA, while a short transcript (terminating in the C2) can only be translated to produce RepA. Control of Rep/RepA levels may also be achieved at the level of transcription, depending on which of three TATA boxes the transcripts are initiated from. For example, transcripts initiated from the two TATA boxes closest to the Rep/RepA start codon are predominantly of the shorter type that are capable of producing only RepA (M.I. Boulton, personal communication).

Regulation of the infection cycle may also be achieved through control of the relative levels of CP and MP expression. Enhancement of the expression of CP relative to MP is achieved in three ways: (1) activation of the *CP* promoter by RepA; (2) expression of CP from both spliced and unspliced versions of a long and short transcript (with particularly high expression from the short, most abundant transcript), while MP can only be expressed with similar efficiencies from the two unspliced transcripts (Wright *et al.*, 1997); and (3) splicing of the *MP* transcript, which not only has the effect of decreasing the amount of transcript available for MP expression, but splicing of the *MP* intron may also greatly enhance CP expression (Wright *et al.*, 1997).

The cell-cycle, phase specific expression of MSV Rep and RepA may provide a clue to the mechanism by which these proteins could control the viral life cycle. Although the experiments of Nikovics *et al.* (2001) could not identify differential expression of Rep and RepA, it is likely that regulation of both proteins' expression at different stages of the infection process (and correspondingly at different phases of the host cell cycle) plays a vital role in the co-ordination of the viral life cycle. RepA appears to be required at all stages of the infection process, from preparing the cellular environment early on for replication, to down-regulating expression of Rep

(thus influencing replication levels) later in the cycle, to activating the expression of CP during the later stages of infection. Mechanisms regulating RepA expression or activity may operate at the transcriptional, post-transcriptional (intron splicing) or post translational stage. The latter may include post-translational processing, as well as biological aspects such as the aggregation state of RepA. Thus, it is clear that MSV has evolved sophisticated mechanisms to ensure tight regulation of both V and C sense expression, possibly co-ordinated by RepA.

Late in the replication cycle, ssDNA is removed from the replication pool and accumulated. It is likely that CP plays a role in this, possibly binding to plus strand DNA released during RCR, arresting the synthesis of new RFI DNAs (Donson *et al.*, 1984). An MSV mutant unable to produce the 13 C-terminal amino acids of the CP fails to accumulate ssDNA (Boulton *et al.*, 1989). However, it is unknown whether this is due to lack of ssDNA sequestration within capsids or whether the mutants lack a specific genetic switch that shifts the infection process from replication to ssDNA accumulation (Palmer and Rybicki, 1998). Interestingly, an MSV mutant containing a point mutation within the CP is able to accumulate ssDNA even though it is unable to form geminate particles (Liu *et al.*, 2001). Although this mutant is competent for all other known functions of the CP (including binding ss and ds DNA, and interaction with the MP), it is incapable of systemic infection. This suggests that (a) encapsidation is not necessary for the accumulation of ssDNA, and (b) encapsidation is necessary for long distance movement of MSV in maize.

After ssDNA has accumulated (but probably before encapsidation is complete), it may need to be moved into neighbouring cells. In MSV this appears to involve both CP and MP (Boulton *et al.*, 1989). Although cell-to-cell movement of both ssDNA and dsDNA is believed to occur in some geminivirus species, ssDNA appears to be the most common form in which geminiviruses move their genomes. MSV CP can bind and transport to the nucleus both ssDNA and dsDNA (Liu *et al.*, 1997), suggesting that it may be the functional equivalent of the bipartite begomovirus nuclear shuttle protein (NSP). The NSP, which also localizes to the nucleus and binds both ss and dsDNA, is believed to mediate the movement of ssDNA into and out of nuclei (Ward and Lazarowitz, 1999). However, the ability of the MSV CP to act as a nuclear shuttle protein is unproven. The reason that the CP binds both ss and dsDNA is unclear. One suggestion is that transport of dsDNA from the nucleus of one cell to that of an adjacent cell might provide a convenient transcription template for immediate early gene expression (Boulton, 2002).

The product of the MSV V1 ORF (MP gene) appears to function as a classical MP: it promotes cell-to-cell movement of viral DNA, it localizes to the cell walls and plasmodesmata in infected maize leaves (Dickinson *et al.*, 1996), and it appears to be capable of modifying plasmodesmatal exlusion limits (Kotlizky *et al.*, 2000). MP also appears to interact with a CP:DNA complex to prevent nuclear import of the viral DNA. The experiments of Kotlizky *et al.* (2000) indicate that MP is able to redirect a proportion of CP:DNA complexes from the nucleus to the cell periphery. A possible model for the role of the CP and MP in the late stage of viral infection is that CP, transports viral DNA to the nucleus for replication and transcription, and then accumulates in the nucleus to eventually encapsidate ssDNA, thus sequestering it from the replication pool. When MP interacts with CP:DNA complexes that form prior to encapsidation, the complexes are targeted to the cell periphery. The CP:DNA complex can then be transported through the plasmodesmata and, following release of the MP, be directed to the nucleus of an adjacent cell to begin another round of replication (Boulton, 2002). The reversal in the directionality of transport (exiting rather than entering the nucleus) may be achieved in a number of ways. For example, interaction of MP with CP may mask the CP NLS, or there may be an as yet undiscovered nuclear export signal on the CP. Alternatively, the balance between nuclear import and export may be regulated by post-translational modification of the CP (Boulton, 2002).

To establish a systemic infection a plant virus moves in two phases: cell-to-cell movement via plasmodesmatal connections, and long distance movement as part of the flow of photoassimilates in the plant vascular system, usually the phloem. Although it is likely that intracellular transport and cell-to-cell movement of MSV DNA is in the form of a nucleprotein complex, it is not clear whether systemic movement relies on normal cell-to-cell movement to deliver genomic DNA into the phloem, or whether viral DNA is specifically packaged for long distance transport. The data of Liu *et al.* (2001) certainly suggest that encapsidation is required for systemic infection.

The mechanisms involved in the encapsidation of geminiviruses are not well understood. There is apparently no encapsidation signal within geminivirus genomes, so it is possible that any circular ssDNA of approximately the right size can be encapsidated (Mansoor *et al.*, 1999; Saunders and Stanley, 1999). MSV plus strands, each with an attached ~80 nt primer molecule (Donson *et al.*, 1984), were shown by Pinner *et al.* (1993) to be packaged into particles that congregate to form large paracrystalline nuclear inclusions. Recently, the structure of MSV particles was resolved by cryo-electron microscopy and 3-D image reconstruction. This showed the MSV particles to consist of two incomplete icosohedra containing 110 copies of the CP arranged with 52-point

symmetry (Zhang *et al.*, 2001). Modelling of the CP revealed that it consists of an eight-stranded antiparallel β-barrel motif, and that the N terminus consists of an α helix containing the putative MSV DNA binding domain. The authors suggest that this region is important for maintaining the geminate particle architecture through interactions with the viral genome (Zhang *et al.*, 2001; Boulton, 2002).

The coat protein is the sole genetic determinant of vector specificity (Briddon *et al.*, 1990). Thus, virus transmission occurs when leafhoppers feed on symptomatic tissues and pick up encapsidated ss MSV DNA, possibly undergoing long distance movement in the phloem. The particles attach at and are transported across the leafhopper hindgut wall, enter the hemocoel and are transported to the salivary glands, thus completing the acquisition process. The leafhopper feeds from the mesophyll and phloem of a healthy plant and delivers the virions to a permissive cell to begin the infection cycle again.

1.3. THE EPIDEMIOLOGY AND CONTROL OF MAIZE STREAK DISEASE

Just as the development of molecular biology techniques in the early 1980s, starting with the cloning and sequencing of geminivirus genomes, led to an intensive era of research on geminivirus molecular biology, the 1990s saw the rapid development of a new age of biotechnology with the potential to revolutionize an important area of research: this was the control of geminivirus diseases. Thanks to the development of genetic engineering, whereby a gene with a desired trait can be transferred into a particular genetic background, the control or even elimination of geminivirus-induced diseases in economically important crops is now viable. However, this cannot be achieved using biotechnology alone. It needs an integrated approach involving, among others, scientists, plant breeders, seed companies, and farmers. For the approach to succeed, a detailed knowledge of the epidemiology of the disease as well as the molecular biology of the viral genome is needed. This section focuses on the epidemiology of maize streak disease (MSD), as well as current and future efforts to achieve its control in maize.

1.3.1 The Epidemiology of Maize Streak Disease

Although geminiviruses can be linked to plant diseases reported over a hundred years ago, it is really only since the early 1990s that these viruses have emerged as a group of serious pathogens that are devastating crops worldwide (Moffat, 1999). Economically important plants as diverse as

tomato, cotton, cassava, wheat, maize, sugarcane, bean, tobacco, beet and horseradish are infected by geminiviruses, sometimes resulting in the destruction of the entire crop. Serious outbreaks of geminivirus-induced disease have occurred in Africa, India, Pakistan, southern Europe, South and Central America, the Caribbean and the USA (Moffat, 1999).

Maize (*Zea mays* L.) was first introduced to Africa in Ghana by Portuguese traders in the 16th century (Gorter, 1953), and has become Africa's most important staple food crop, increasingly replacing traditional food crops such as sorghum and millet. Whereas the worldwide average maize yield is ~4 tons hectare^{-1} (with highly industrialized nations averaging ~8 tons hectare^{-1}), the average maize yield in Africa is the lowest in the world at ~1.7 tons hectare^{-1} (Wambugu and Wafula, 1999). Maize pathogens coupled with outmoded agricultural practices are the main reasons for poor yields in Africa. Of the many pathogens infecting maize, MSV is considered the most important and widespread (Thottappilly *et al.*, 1993). Indigenous to sub-Saharan Africa and the neighbouring Indian Ocean islands of Madagascar, Mauritius and La Réunion (Bosque-Pérez, 2000), MSV can result in maize yield losses of up to 100% (Wambugu and Wafula, 1999).

The symptoms of MSD first appear on the lowest exposed portion of the maize leaf as roughly circular spots (Bock *et al.*, 1974), which develop into chlorotic streaks as the leaf expands. Symptoms on mature leaves range from narrow veinal streaks a fraction of a millimeter in width to complete leaf chlorosis. The chlorosis is caused by the failure of chloroplasts to develop in the tissue surrounding the vascular bundles, which results in reduced photosynthesis and increased respiration, leading to a reduction in leaf length and plant height. Symptom severity depends on the MSV strain, the host genotype, and the age of the plant at the time of infection. Highly sensitive varieties or plants infected at an early stage become severely stunted, producing undersized, deformed cobs or giving no yield at all. In the most severe cases, chlorosis of the entire leaf is followed by progressive necrosis and plant death (Bosque-Pérez, 2000).

Many factors determine the severity and frequency of an outbreak of MSD. In particular, the disease is dependent on the complex relationship between the host plant/s (maize, sugarcane and over 80 grass species), a large, unknown number of MSV strains (Konate and Traore, 1992), and the leafhopper vector (*Cicadulina* spp.), nine species of which are able to transmit MSV (Bosque-Pérez, 2000). Environmental factors that have an influence on the leafhopper population also play an important role in MSD epidemiology. For example, MSD outbreaks are often associated with drought conditions followed by irregular rains at the beginning of growing seasons (Efron *et al.*,

1989), as in the savanna regions of West Africa in 1983 and 1984 (Rossel and Thottappilly, 1985), or in Kenya in 1988-89 (Njuguna *et al.*, 1990). The relative abundance of different *Cicadulina* species with differing abilities to transmit the virus in different parts of Africa, is influenced by altitude, temperature, and rainfall (Dabrowski *et al.*, 1987). In addition, late rainfall favours the development of leafhopper nymphs during the winter (Stanley *et al.*, 1998). The interplay of all these factors makes MSD rather erratic, being devastating some years and insignificant in others (Efron *et al.*, 1989).

Although MSV is transmitted by at least nine *Cicadulina* species, *C. mbila* (Naudè) is the species most often implicated in MSD outbreaks (Dabrowski, 1987). This is because *C. mbila* is the most widely distributed species, covering the entire continent of Africa, as well as parts of Asia, Australia, the Indian and Pacific Islands and northern parts of South America (Rose, 1978). Additionally, a larger proportion of *C. mbila* populations have the ability to transmit MSV, compared with other *Cicadulina* species (Storey, 1928, 1933; Markham *et al.*, 1984). This is partly due to the proportion of *C. mbila* females, which are better transmitters, being 2-3 times higher than in other species (Wambugu and Wafula, 1999).

Leafhoppers do not breed on maize; their favoured hosts are annual grasses. Grasses are most probably MSV's natural reservoir sources, from which they "emerged" into maize when the crop was introduced into Africa (Rybicki and Pieterson, 1999). *C. mbila* can feed on more than 138 grass species, ~70% of which are potential MSV hosts (Konate and Traore, 1992). Thus, grass species have a great influence on MSD epidemiology. For example, the species composition and age distribution of grasses in an area may strongly influence the amount of MSV inoculum available for transfer in that area.

When feeding on an infected plant, a leafhopper can acquire the virus by feeding directly on chlorotic lesions and on the phloem where the virus circulates. The length of time required for a leafhopper to acquire MSV while feeding is between 5 and 20 seconds, while the minimum transmission time is between 5 and 10 minutes (Storey, 1925, 1938). The leafhopper becomes viruliferous within 30 hours following acquisition of the virus, although at 30°C the latent period can be as short as 6 to 12 hours (Storey, 1938). While there is no evidence of transovarial transmission of MSV in *Cicadulina* (Storey, 1928), leafhopper nymphs are able to acquire MSV soon after emerging from the egg, and retain the ability to acquire and transmit the virus throughout their lifetime. However, virus titres within a leafhopper decrease over the lifetime of

the insect, indicating that MSV replication does not occur within leafhoppers (Reynaud and Peterschmitt, 1992).

The mechanisms by which MSV particles are transported into the leafhopper's haemocoel and then into its salivary glands are unknown, but they are likely to play a role in determining the ability of the leafhopper to transmit the virus, which is an inherited, dominant sex-linked characteristic (Storey, 1932). Thus, a leafhopper may acquire the virus from the phloem of infected plants, but if the virus cannot, for example, attach to the leafhopper hindgut wall or be transported across the wall to the haemocoel, it cannot be transmitted to another plant. In addition, studies of feeding activities of *C. mbila* on different hosts by Mesfin *et al.* (1995) have revealed vector preferences for certain hosts, which may also play an important role in virus transmission from one host to another. For example, inoculation of MSV into healthy maize occurs when insects salivate into the phloem tissue (Kimmins and Bosque-Pérez, 1996); the time taken to reach the phloem and transmit MSV may take as long as one to three hours from initial access (Bosque-Pérez, 2000). Since leafhoppers making brief probes may not reach the phloem, transmission of MSV thus requires long-duration probing. In this way, the feeding behaviour of *Cicadulina* on a maize genotype has an influence on the resistance of the variety to MSV. On hosts from which the leafhopper does not prefer to feed, the ability to transmit MSV is reduced due to shorter probing times (Bosque-Pérez, 2000).

The flight behaviour of leafhopper populations is another factor that determines the incidence and severity of MSD. Leafhoppers move either within a crop, or migrate from maturing crops or from perennial alternative hosts into younger crop or grass hosts (Bosque- Pérez, 2000). Distinct long and short distance flight morphs have been detected amongst *Cicadulina* populations in Zimbabwe (Rose, 1972), which determine the distance that MSV spreads from the source of inoculum. The long flight morphs, believed to be the migratory form, may play an important part in the long distance spread of virulent MSV variants (Rose, 1978). In turn, migratory movement is probably influenced by environmental conditions such as rainfall and temperature, as well as wind, since *Cicadulina* disperse downwind (Rose, 1972). In Zimbabwe and Nigeria, population densities of *Cicadulina* are low at the onset of the rainy season and then rise gradually as host plants become abundant and succulent. Populations decline sharply in the dry season, probably due to both adverse environmental conditions and an absence of hosts on which to feed (Bosque-Pérez, 2000). Other factors influencing leafhopper flight behaviour are the season, time of day,

gender (females fly further than males), presence of mature ova in the females, age of the leafhopper, and condition of the plants on which the leafhoppers feed (Rose, 1972).

In summary, outbreaks of MSD appear to occur only when favourable weather conditions allow leafhopper survival and population build-up, and where MSV infects both grass and maize hosts. Certain agricultural practices increase the chances of epidemics occurring. During the past 15 years the area of maize cultivation in Africa has greatly increased, and the crop is now often grown as a monoculture (Bosque- Pérez, 2000). In addition, emphasis in the past has been on breeding to improve yields, resulting in high yielding varieties that are also highly susceptible to MSV (Wambugu and Wafula, 1999). Indeed, the introduction of new susceptible genotypes and the increased area under maize are believed to be two of the main factors leading to increased MSV and *Cicadulina* occurrences in Africa (Bosque- Pérez, 2000). Growing crops year-round under irrigation has also contributed to the increased incidence of MSD. For example, in Zimbabwe cereal crops such as wheat are cultivated in the dry season, serving as a host for leafhoppers that later move to early-planted maize (Rose, 1973). Similarly, in many parts of Africa (especially West Africa) maize is grown all year round. Both instances provide year-round food and suitable oviposition sites for leafhoppers. The continuous presence of these crops and the grasses associated with them has serious implications for MSD epidemiology (Bosque-Pérez, 2000). The following section discusses the combined strategies required for the effective, durable control of MSD. These include changing cultural practices, and creating resistant cultivars, either by classical cross-breeding or through genetic engineering, or a combination of both.

1.3.2 Strategies for the Control of Maize Streak Disease

Historically, attempts at control of MSD have focused on evasive measures and the breeding of maize for naturally occurring resistance. Evasive measures include control of the leafhopper population using insecticides, and various cultural practices. Insecticides have been largely unsuccessful due to the need for frequent spraying, which is not only ecologically undesirable, but is expensive and can lead to insecticide resistance within leafhopper populations (Stanley *et al.*, 1998). Cultural practices suggested for control include barriers of bare ground between early- and late-planted maize fields to reduce leafhopper movement and subsequent spread of MSV (Bosque- Pérez, 2000), avoiding maize plantings downwind from older cereal crops, and the use of crop rotations that will minimize invasion by viruliferous leafhoppers (Rose, 1978). Of the

traditional control measures, resistance breeding is perceived as the most practical solution for the control of MSV.

1.3.2.1 Classical cross-breeding for MSV resistance

Resistance in maize was noted as early as 1931 in South Africa, in the variety "Peruvian Yellow" (Fielding, 1933), and several other varieties have since been found to have varying degrees of resistance (or tolerance). Resistance usually manifests itself as reduced symptom severity combined with low virus titres, leading to low virus incidence in the field. Resistant varieties are therefore much poorer sources of inoculum during secondary disease spread (Rodier *et al.*, 1995). In some cases resistant varieties yield well even when infected (Bosque- Pérez, 2000).

Several national programmes in Africa (including South Africa, Zimbabwe, Nigeria, Kenya, and La Réunion) are breeding for resistance to MSV, combining the resistance with other desirable characteristics (Rybicki and Pieterson, 1999). For example, Nigeria's International Institute of Tropical Agriculture (IITA), in cooperation with Zimbabwe's International Maize and Wheat Improvement Center (CIMMYT) and National Programs in Africa, have incorporated MSV resistance into high-yielding varieties and varieties traditionally grown in various African countries (Efron *et al.*, 1989). In addition, the Pannar seed company of South Africa has developed and released MSV-resistant hybrids in several African countries (Bosque-Pérez, 2000).

Despite these extensive efforts to breed resistant maize, there has been only limited success in the field (Stanley *et al.*, 1998). Although in certain countries such as Nigeria there are examples of MSV-resistant varieties being largely unaffected by MSV over many years (Bosque-Pérez, 2000), there are also a number of reports of severe infection of so-called MSV-tolerant maize (Stanley *et al.*, 1998). This occurs particularly when varieties are grown under environmental conditions different from those in which the plants were selected. In addition, some maize varieties known to be resistant elsewhere are susceptible to certain viral strains/isolates, as reported in La Réunion (Rodier *et al.*, cited by Bosque-Pérez, 2000). In Kenya, despite having one of Africa's leading national agricultural research institutions, the Kenya Agricultural Research Institute (KARI), in which plant pathologists, entomologists and plant breeders have worked on MSV for two decades, there has recently been an escalation in MSD incidence (Wambugu and Wafula, 1999). For example, whereas in 1978 a survey revealed low and sporadic incidence of MSD in maize farms in 34 districts in Kenya, in 1988 infection rates as high as 70-80% were reported in the Central Highlands, and in 1998 a survey of ten farms in southwest Kenya showed a MSD

incidence of 80-100% (Wambugu and Wafula, 1999). The increase of MSD incidence over the years could be due to several factors. Possibilities include the emergence of new, more virulent MSV strains, an increase in MSV-susceptible maize varieties grown by farmers, and the breakdown of resistance in the "resistant" varieties. New MSV strains in a particular area could arise from mutation or recombination of the viral genome, or migration via the leafhopper, and have the potential to break the resistance developed against less virulent strains. Natural inherited virus resistance in plants is thought to be due to one or more viral genes (e.g. the genes encoding the coat protein, replicase, or movement protein) encoding an avirulence factor that elicits resistance controlled by a cognate dominant host gene. Some resistance-breaking virus variants have merely a single nucleotide replacement in their avirulence gene (Harrison, 2002). The probablility of a resistance-breaking variant appearing depends on the type of resistance and the number of resistance genes to be overcome. The fact that the maize genes involved in MSV resistance are poorly understood makes it difficult to create a variety that is resistant to a broad range of viral strains, especially when the virus frequently mutates and recombines (D. Martin, pers. comm.). Thus, it remains to be seen whether the MSV resistance of commercially available maize varieties is durable and effective over a long period.

Even if long-term resistance can be achieved by classical cross-breeding techniques, there are drawbacks to conventional breeding programmes, which often prove to be difficult and time-consuming. One of the major constraints facing breeders is the need to maintain crop quality and yield while introducing resistance traits (Frischmuth and Stanley, 1993). Genetic resistance is not usually readily available in the desired cultivar, and as mentioned previously, is not necessarily effective against all the strains of a given virus. Resistance genes derived from other plant lines or species are difficult to transfer because they are usually associated with undesirable characteristics, or they can be polygenic in nature (Hemenway *et al.*, 1989).

Modern gene transfer techniques are faster and more precise than classical plant breeding (Moffat, 1999). In recent years the transformation of many plant species has become routine, and genetic engineering is becoming an important part of plant breeding programmes. There are two main advantages of genetic engineering. These are: (1) the ability to transfer single genes directly without linkage to undesired genes, and (2) the ability to construct novel genes that are unlikely to exist in nature (Gadani *et al.*, 1990). This makes it easier to design a resistance strategy that is more likely to succeed against a broad range of virus strains, and that is less likely to be broken than natural resistance. Using genetic engineering, a number of promising strategies (discussed in

the following section) have been developed to introduce geminivirus resistance into economically important crops. Most work has focused on begomovirus diseases due to the recalcitrance of cereals to transformation, but the principles can be readily applied to mastrevirus diseases such as MSD.

1.3.2.2 Genetic engineering of plants for geminivirus resistance

The first approach to genetically engineering plants for virus resistance was to mimic the natural phenomenon of "cross protection", first observed over 70 years ago by McKinney (1929), who showed that plants already infected with a virus are normally protected against infection by a related strain of the virus. Thus, plants infected with a mild strain could be protected against infection by severe isolates or strains of that virus. There were a number of reasons for the limited use of this form of protection, including the impracticality of inoculating plants with a live virus on a large scale, and the possibility that the mild strain might mutate to a more virulent form. In 1985, Sanford and Johnson used the principle of cross protection to develop the concept of parasite- or pathogen-derived resistance (PDR). A simple but elegant concept, PDR is the process whereby resistance to a pathogen may be engineered by transforming a susceptible plant with a gene derived from the pathogen itself. In general, PDR is thought to operate through the expression of the viral gene product at either an inappropriate time, in inappropriate amounts, or in an inappropriate form during the infection cycle, thereby perturbing the ability of the pathogen to sustain an infection (Lomonossoff, 1995). The first demonstration of virus-derived resistance in transgenic plants made use of the coat protein (CP) gene of *Tobacco mosaic virus* (TMV; Powell-Abel *et al.*, 1986). CP-mediated protection, which has subsequently been used successfully against a number of viruses, is thought to operate through the inhibition of virion disassembly in the initially infected cells (Baulcombe, 1996).

Other viral genes shown to be capable of conferring PDR are the replicase gene, first demonstrated with TMV and subsequently found to be effective with numerous other viruses, and the movement protein gene from tobamoviruses, bromoviruses and potexviruses (Lomonossoff, 1995). The PDR concept has been applied mainly to viruses with positive-strand RNA genomes, and most examples of CP-mediated resistance are based on the transgenic expression of wild-type genes. Although there is one report of wild type CP-mediated resistance against a geminivirus (TYLCV; Kunik *et al.*, 1994), this is not considered to be the best strategy to engineer resistance against geminiviruses. However, other effective strategies are based on variations of the PDR concept, including the transformation of susceptible plants with viral genes containing dominant

negative mutations (Herskowitz, 1987), defective interfering DNA derived from the viral genome, and antisense viral RNA. There are also examples of PDR occurring as a result of gene silencing (Baulcombe, 1996). The following are examples of the successful application of engineered resistance strategies against geminiviruses.

(1) Defective Interfering DNA

Subgenomic DNA molecules that are related to and dependent on the parent virus for their proliferation have been found associated with geminivirus infection. Because some subgenomic DNAs have the ability to delay and attenuate infection symptoms, they are referred to as defective interfering (DI) DNA (Frischmuth *et al.*, 1997). Transgenic *N. benthamiana* plants containing tandemly repeated, integrated copies of ACMV (Stanley *et al.*, 1990; Frischmuth and Stanley, 1991) or BCTV (Frischmuth and Stanley, 1994) DI-DNAs exhibit ameliorated symptoms when challenge-inoculated with the homologous parent virus. The mechanism of resistance is thought to occur through the mobilization of extrachromosomal copies of DI DNA following virus infection, through Rep-mediated release. Subsequent episomal amplification of the DI DNA occurs at the expense of the genomic viral DNA, resulting in reduced virus amplification and symptom amelioration (Stenger, 1994). This strategy is limited, since it relies on the ability of the infecting virus to replicate the DI DNA. Thus, plants show resistance only to closely related strains of virus from which the DI DNA is derived (Stanley *et al.*, 1998).

(2) Antisense RNA

RNA molecules capable of annealing to a given mRNA are a means of natural and artificial gene regulation by silencing the expression of the corresponding gene (Bendahmane and Gronenborn, 1997). This phenomenon has been successfully exploited to target and selectively suppress the expression of specific genes, in both therapeutics and in the prevention of viral diseases. It is likely that antisense RNAs anneal with sense RNAs to form a double strand complex, which is rapidly degraded or which inhibits the translation of the RNA. The nuclear-replicating geminiviruses are potentially promising targets for antisense RNA-mediated suppression, since the antisense sequences do not have to be directed to the cytoplasm as they do with cytoplasmic RNA viruses (Frischmuth and Stanley, 1993).

Antisense RNAs are usually targeted at the C sense genes that not only are indispensable for viral replication and required early in the infection cycle, but are also expressed from transcripts of relatively low abundance. These factors make the Rep gene an ideal target for obtaining

geminivirus resistance by suppressing its expression. This approach has met with some success in TYLCV- and TGMV-susceptible plants. In transgenic *N. tabacum* plants expressing the TGMV Rep antisense RNA, symptoms were less severe compared with non-transformed control plants when challenged with TGMV (Day *et al.*, 1991). In addition, most antisense lines contained fewer symptomatic plants than control plants and transgenic lines expressing Rep sense RNA. Interestingly, Bejarano and Lichtenstein (1994) showed that TGMV antisense RNA was also effective against BCTV, but not ACMV. The authors concluded that a minimal contiguous complementarity between the antisense RNA and the mRNA target is required for efficient suppression. *N. benthamiana* plants that are less susceptible to TYLCV infection have also been produced by expressing TYLCV Rep antisense RNA (Bendahmane and Gronenborn, 1997). Some of the resistant lines were symptomless, and the replication of TYLCV almost completely suppressed. These reports suggest that antisense RNA is a promising resistance strategy, but the effectiveness of this approach against other geminiviruses depends on the level of homology between the target sequence and the antisense RNA (Stanley *et al.*, 1997).

(3) Expression of Viral Proteins

The expression of geminivirus proteins or their derivatives in transgenic plants has conferred resistance against a number of geminiviruses. An ideal candidate for interference with virus infection is Rep, which is multifunctional, required early in the infection cycle and expressed at low levels from the Rep promoter. However, a novel resistance strategy based on the suppression of virus movement was developed for bipartite geminiviruses by von Arnim and Stanley (1992a, b). They found that the TGMV MP, which cannot complement movement of ACMV, specifically inhibits ACMV systemic spread. It is likely that the ACMV MP functions as a multimeric component of a movement complex that is disrupted by the inclusion of one or a limited number of copies of the inhibitory protein (Frischmuth and Stanley, 1993). A variation of this strategy was employed by Duan *et al.* (1997), who transformed *N. tabacum* with a mutated MP gene from *Tomato mottle virus* (TMoV). Transgenic plants expressing the defective MP showed resistance to both TMoV and *Cabbage leaf curl virus* (CabLCV). The degree of resistance correlated with the level of expression, suggesting that the defective protein functions as a dominant negative mutant of a movement function. In a similar experiment, Hou *et al.* (2000) found that tomato plants transformed with a mutated *Bean dwarf mosaic virus* (BDMV) MP showed a delay in ToMV infection compared with non-transformed plants. The fact that the resistance in the cases of both Duan *et al.* (1997) and Hou *et al.* (2000) extended to heterologous geminiviruses, suggests that this approach may result in broader spectrum resistance than strategies that target

viral DNA replication (Frischmuth and Stanley, 1993). However, expression of wild type MP, which is a pathogenicity determinant, can have deleterious effects of various aspects of plant development (Covey and Al-Kaff, 2000; Hou *et al.*, 2000), necessitating the use of a defective *MP* transgene in order to regenerate phenotypically normal plants.

One of the first steps in virus multiplication is the replication of the viral genome. Therefore, blocking this phase should be one of the most efficient ways to protect plants from geminivirus infection. Noris *et al.* (1996) and Brunetti *et al.* (1997) showed that expression of a truncated TYLCV Rep gene (T-Rep), in transgenic *N. benthamiana* and tomato respectively, interfered with TYLCV infection. However, plants expressing TYLCV T-Rep were not protected against ToLCV, suggesting that the resistance mechanism is specific. Brunetti *et al.* (2001) determined that T-Rep acts as a *trans*-dominant negative mutant, inhibiting both viral transcription and replication. They proposed a model of TYLCV resistance conferred by T-Rep. Initially, T-Rep inhibits but does not abolish Rep transcription by recognizing and binding to the cognate transcriptionally active RF DNA. The limited amount of newly synthesized Rep cannot properly synthesise the viral plus strand since it is out-competed by T-Rep for utilization of the required sequence; as a result, virus replication is inhibited.

Hong and Stanley (1996) used a similar approach in conferring resistance to ACMV in *N. benthamiana*. They found that transient expression of ACMV Rep or a truncated N-terminal portion of the protein caused a significant reduction in the level of viral DNA in *N. tabacum* protoplasts. Subsequent transformation of *N. benthamiana* with the full-length Rep gene resulted in plants that were less susceptible to infection by ACMV. None of the transformed lines showed resistance to TGMV or BCTV, again demonstrating the specific nature of the resistance mechanism. The authors suggest that the transgenic Rep protein, which is highly expressed by the CaMV 35S promoter, could affect viral DNA replication by disturbing the equilibrium between monomeric and multimeric forms of the viral Rep complexes. Resistance to ACMV has also been achieved in *N. benthamiana* by the high-level expression of the Rep gene containing a mutation of the NTP-binding domain (Sangaré *et al.*, 1999).

Transient expression of a truncated ToLCV Rep protein was found to inhibit homologous viral DNA accumulation in tobacco protoplasts and in *N. benthamiana* plants (Chatterji *et al.*, 2001). This protein, which contains the N-terminal 160 amino acids of Rep and therefore the sites for DNA cleavage, DNA binding, and protein oligomerization, was shown to interfere with DNA

binding and oligomerization activities during virus infection. Surprisingly, the truncated protein also reduced accumulation of *Pepper huasteco yellow vein virus* (PHYVV) and *Potato yellow mosaic virus* (PYMV), although to a lesser extent than ToLCV. The truncated ToLCV Rep formed oligomers with the Rep proteins of the heterologous geminiviruses, which suggests that it may function as a dominant negative mutant, interfering with one or more of the multiple functions of the wild type Rep oligomers.

As well as published reports, there have been several patent applications relating to begomovirus resistance in transgenic plants conferred by Rep genes mutated in one or more of the conserved motifs I-IV (Hanson *et al.*, 1998; Gronenborn, 2000, Stout *et al.*, 2001; Hanley-Bowdoin *et al.*, 2002a) as well as a mutant form of the begomoviral accessory replication protein, REn (Hanley-Bowdoin *et al.*, 2002b). In general, the mutant Rep (or REn) in all these cases may interfere with the replication activity of the wild type protein expressed by infecting begomoviruses, in this way behaving as dominant negative mutants.

(4) Virus-induced Cell Death

Using a mechanism involving a viral-activated antiviral protein, Hong *et al.* (1996) conferred resistance to ACMV in *N. benthamiana* by the expression of a ribosome inactivating protein (RIP) from the ACMV V sense promoter. The RIP used was dianthin, a potent plant cytotoxin, which rapidly kills infected cells, thereby containing the infection. This resistance mechanism depends on the transactivation property of TrAP, which induces the expression of the ACMV CP. Thus, expression of dianthin under the control of the *CP* promoter should be activated specifically once TrAP is expressed in virus-infected cells, avoiding constitutive expression of the RIP and ensuring transgene expression is localized to virus-infected cells. Hong *et al.* (1996) found that plants containing the dianthin gene were less susceptible to virus infection and accumulated only low levels of viral DNA. A drawback to this method is that a low level of constitutive expression from the *CP* promoter can occur in the absence of TrAP. Another is that the resistance was confined to ACMV isolates, implying that TrAP activity is virus-specific (Stanley *et al.*, 1998)

(5) Gene Silencing

The examples of PDR discussed above are those in which expression of the transgene leads to the resistance phenotype. However, in some types of PDR gene silencing is involved in and is responsible for the resistance mechanism (Baulcombe, 1996). Gene silencing is a cellular

mechanism that targets specific nucleic acid sequences for down-regulation or degradation (Covey and Al-Kaff, 2000). First observed in plants containing transgenes, gene silencing can operate both at the transcriptional and post-transcriptional levels. Transgene-associated post-transcriptional gene silencing (PTGS) targets the transgene RNA and RNA from homologous endogenous genes for degradation in the cytoplasm (Covey and Al-Kaff, 2000). Experiments with transgenic plants expressing viral sequences led to the discovery of virus-associated PTGS, whereby, upon infection by the virus, the virus-derived transgene leads to the suppression of expression of both the transgene and the homologous viral gene. If the virus gene is an essential component of its lifecycle, for example the Rep gene in the case of geminiviruses, PTGS of the gene leads to virus inactivation. Although most cases of genetically engineered geminivirus resistance have been associated with high levels of transgene expression and transgene product, there are cases of transgene-mediated resistance to geminiviruses that may have been mediated at the RNA level. For example, resistance to TMoV was found in transgenic plants expressing the TMoV CP gene with a 5' deletion of 30 bp (Sinisterra *et al.*, 1999, cited by Covey and Al-Kaff, 2000), but the resistance did not appear to be the result of the classical CP-mediated protection. In those plants where a resistance phenotype was observed, transgene RNA, but no transgene protein, was detected. The authors concluded that this resistance may have been mediated at the RNA level, although specific characteristics typical of PTGS, such as co-reduced levels of transgene and viral RNA, were not determined. Although there is not much evidence for gene silencing elicited by transgenic expression of individual geminiviral genes or fragments, geminivirus replicons or vectors do have the ability to induce PTGS. For example, when Kjemtrup *et al.* (1998) used TGMV as a vector to carry a foreign sequence in place of the CP gene, it triggered gene silencing of both the vector sequence and the homologous endogene when introduced into *N. benthamiana*. Gene silencing elicited from an integrated geminiviral vector releasing the mastrevirus TYDV (an autonomously replicating multicopy plant episome, or MPE) has also been demonstrated (Atkinson *et al.*, 1998).

Gene silencing appears to be one of several host mechanisms operating against viruses. Since it is clear that geminiviruses can elicit PTGS targeted against viral RNAs, it could be possible to trigger PTGS targeted against an essential geminiviral RNA, leading to viral inactivation. Thus, virus-derived sequences could be used as transgenes to provide resistance based on PTGS. However, there are some drawbacks to this resistance mechanism. Because gene silencing is homology-dependent, there is likely to be only a limited amount of sequence non-identity between the transgene and the challenge virus that can be tolerated by the silencing mechanism. It

is also becoming clear that some viruses, including geminiviruses, contain anti-silencing genetic determinants, which lead to the suppression of gene silencing (Kasshau and Carrington, 1998; Voinnet *et al.*, 1999). One such determinant is the ACMV AC2 gene, encoding TrAP. The anti-silencing determinants associated with AC2 were assayed in transgenic plants containing a green fluorescent protein (GFP) transgene that had been systemically silenced by PTGS. Subsequent infection by a potato virus X (PVX) vector expressing the ACMV AC2 gene led to suppression of silencing of the GFP gene (Voinnet *et al.*, 1999).

It is clear from the above example that there is a more balanced interaction between plants and viruses than previously thought, involving host defense and viral counter-defensive strategies. The best, most durable viral resistance is therefore likely to have the following characteristics: (1) protection against a wide range of viral strains; (2) multiple mutations (preferably more than two) in the viral genome required to overcome host resistance, and (3) a resistance mechanism that confines the virus to the inoculated cell (Harrison, 2002). Examples of the latter include blocking replication or cell-to-cell movement of the virus, or inhibiting the expression of a protein required early in the infection cycle by an RNA-mediated resistance mechanism such as antisense RNA or PTGS. Development of this sort of resistance, together with sound crop management as discussed at the beginning of this section, may be the only long-term means of successfully combating MSD.

1.4 PROJECT AIMS

The main aim of this project was the development of a genetic engineering approach, to ultimately result in the development of MSV-resistant maize. The strategy chosen was that of pathogen-derived resistance (PDR), using the multifunctional viral Rep gene as a target for mutagenesis and truncation. Rep constructs had previously been made containing deleterious mutations in several conserved motifs and had been shown in transient expression assays to be effective in inhibiting MSV replication. In this study, the first aim was to truncate these mutants as well as the wild type Rep gene, in order to remove key motifs involved in viral replication and in the interaction with host factors that could potentially interfere with plant development. A three-step approach was designed to test the effects of these constructs on viral replication. First, truncated and mutated Rep constructs would be co-bombarded with a widespread representative MSV strain (MSV-Kom) into black Mexican sweetcorn (BMS) suspension cells, and the effects of the constructs on viral replication determined by developing a quantitative PCR assay. Second,

Rep constructs that inhibited viral replication in BMS would be used to transform an MSV-sensitive grass, *Digitaria sanguinalis*. Because of the recalcitrance of maize to tissue culture, *D. sanguinalis*, which is easily transformable and regenerable, is useful as a model system to test the effect on viral replication of the transgenic expression of a range of mutated or truncated Rep proteins. To do this, transgenic plants would be challenged with MSV using viruliferous leafhoppers, the vectors of the virus. The third step of the approach, assuming MSV-resistant *D. sanguinalis* was produced, would be to transform the maize cultivar High Type II (Hi-II) with Rep constructs that resulted in the best, most durable resistance in *D. sanguinalis*.

In the process of carrying out the above transient and transgenic assays, some interesting results were obtained with some of the mutated and truncated Rep constructs, which had unexpected effects on viral replication when co-bombarded with MSV-Kom into BMS. As a sideline to the main aim, a particular focus was a Rep gene containing a mutation in the retinoblastoma-related (RBR) protein interaction domain, which contrary to published reports at the time, surprisingly supported high-level viral replication in BMS, while *in planta* one of the nucleotides of the three-nucleotide mutation reverted at an extremely high frequency. The remaining aims of the project were to carry out a study of the dynamics of the mutant and revertant viral populations in maize, and to determine the selective advantage of the single nucleotide reversion.

Chapter 2

Inhibition of *Maize streak virus* (MSV) Replication by Transient and Transgenic Expression of MSV Replication-Associated Protein Mutants

Contents

ABSTRACT

Maize streak virus (MSV), the type member of the *Mastrevirus* genus of the *Geminiviridae*, is the worst viral disease-causing agent affecting maize in Africa, resulting in crop yield losses of up to 100%. To genetically engineer MSV-resistant maize using the pathogen-derived resistance (PDR) strategy, the viral replication-associated (Rep) protein gene was targeted, whose multifunctional products are the only viral proteins essential for replication. *Rep* constructs had previously been made containing deleterious mutations in several conserved motifs. In this study, these mutant *Reps* and the wild type Rep gene were truncated to remove key motifs involved in viral replication. A quantitative PCR assay was developed to determine the effects of the mutant and truncated Reps on viral replication in black Mexican sweetcorn (BMS) suspension cells. The MSV-sensitive grass *Digitaria sanguinalis* was then transformed with *Rep* constructs that inhibited MSV replication in BMS, and transgenic lines were tested for virus resistance. Several plants of a *D. sanguinalis* line transgenic for a mutated full-length Rep gene showed excellent resistance (immunity) to MSV, but the transgene had negative effects on aspects of plant growth and development. Transformation with a mutated/truncated Rep gene, however, resulted in healthy fertile transgenic *D. sanguinalis* plants, many of which showed good MSV resistance in challenge experiments. Resistance phenotypes included a delay in symptom development, a recovery from early symptoms, and an absence of virus symptoms at all stages. Expression of the transgenes in lines that were challenged with MSV was determined by RT-PCR and by histochemical staining for GUS (the transforming plasmid contained the *uidA* gene). The maize cultivar Hi-II was transformed with the mutated/truncated Rep gene, and three transgenic lines produced fertile T_1 offspring. Considering the success in achieving MSV-resistant *D. sanguinalis*, there is good reason to believe that the transgenic maize will too be resistant to MSV.

2.1 INTRODUCTION

Maize is Africa's most important staple food crop and is increasingly replacing traditional food crops such as sorghum and millet. In most of sub-Saharan Africa poor maize yields are usually linked to food shortages and famine. Despite being a crucial staple food crop, the average maize yield per hectare in Africa is the lowest in the world. A major contributing factor to these low yields is the causal agent of maize streak disease (MSD), *Maize streak virus* (MSV). Transmitted by leafhoppers (*Cicadulina* sp.), MSV is indigenous to sub-Saharan Africa and neighbouring Indian Ocean Islands.

Methods of reducing yield losses caused by MSV include the use of insecticides to control leafhopper populations, and the use of MSV-resistant maize genotypes. Insecticides are expensive and are beyond the means of most farmers in Africa. Frequent spraying of insecticides is undesirable as it may lead to insecticide resistance, and is ecologically unfavourable. Conventional breeding programmes are protracted, and there has been limited success in combining economically important traits such as yield with resistance characteristics. Another important drawback to conventional breeding is that the resistance can be circumvented by virus variation.

A more efficient, cost-effective and sustainable solution could be the development of MSV-resistant maize by genetic engineering. Transformation of plants with viral genes can give rise to lines of plants that are resistant to the virus from which the sequence was derived. Although this phenomenon, termed "pathogen derived resistance" (PDR), has been successfully applied for resistance to viruses of the *Begomovirus* genus of the *Geminiviridae*, there are no published reports of any transgenic plants resistant to mastreviruses using this or any other approach. With begomoviruses, coat protein-mediated protection against *Tomato yellow leaf curl virus* (TYLCV) has been achieved in tomato (Kunik *et al.*, 1994); interference with *African cassava mosaic virus* (ACMV) replication by defective-interfering viral genomes in transgenic plants has been reported (Stanley *et al.*, 1990; Frischmuth and Stanley, 1991), and defective movement protein has resulted in resistance to *Tomato mottle virus* (ToMoV, Duan *et al.*, 1997) and *Bean dwarf mosaic virus* (BDMV; Hou *et al.*, 2000). Other approaches for interfering with begomovirus replication have included expression of the antisense RNA of *Tomato golden mosaic virus* (TGMV, Day *et al*; 1991; Bejarano and Lichtenstein, 1994) and TYLCV (Bendahmane and Gronenborn, 1997), and truncations of the viral replication-associated protein (Rep). There are two reports of resistance being achieved (although limited) with the latter approach, against TYLCV (Noris *et al.*, 1996) and ACMV (Hong and Stanley, 1996). In addition, a mutant viral Rep protein was used by Sangare *et al.* (1999) to develop resistance against ACMV in tobacco.

To genetically engineer MSV-resistant plants, the strategy employed was that of virus-derived resistance, by means of dominant negative mutant interference (Herskowitz, 1987) with virus replication. The viral gene chosen for mutagenesis and truncation was *Rep*, whose multifunctional products, Rep and RepA, are the only viral proteins essential for replication.

MSV replicates in the nucleus by a rolling circle replication (RCR) mechanism (Saunders *et al.*, 1991; Stenger *et al.*, 1991). This is initiated by the binding of Rep to the origin of replication, where the protein introduces a sequence-specific nick in the loop of a stem-loop structure. Host replication enzymes then complete the RCR process.

Rep is a product of transcription of two open reading frames (ORFs), *C1* and *C2*. The *C1/C2* transcript has an intron, and depending on whether or not it is spliced, expresses either Rep (from the spliced transcript) or RepA (from the unspliced transcript). Rep and RepA have several distinct domains with diverse biochemical activities (see Fig. 2.1), among them DNA-binding, nicking-joining, transactivation (all activities shared by Rep and RepA), interaction with the host retinoblastoma related (RBR) protein (RepA only), and NTP-binding/ATPase activities (Rep only).

In a previous study from this laboratory, the following conserved motifs of MSV Rep were chosen as targets for PCR-mediated mutagenesis: motif III, which is essential for nicking and closing the DNA during rolling circle replication; the LxCxE motif of the RBR protein-interaction domain, which may enable the virus to create an optimal cellular environment for virus replication; and a motif in the NTP binding domain, which may be necessary for Rep helicase activities. Single, double (two motifs mutated) or triple (all three motifs mutated) Rep mutants were generated by PCR mutagenesis of the Rep gene (T. Mangwende, 2001). In the present study, mutated and wild type *Rep* constructs were truncated to yield N-terminal fragments missing the entire C2 ORF and the C terminus of the RepA gene.

To determine the effects of the mutant and truncated Reps on viral replication, a three-system approach was designed. First, a quantitative PCR assay was developed to accurately compare the effect of the mutant and truncated Reps on viral replication in black Mexican sweetcorn (BMS) suspension cells. Second, an MSV-sensitive grass, *Digitaria sanguinalis*, was transformed with *Rep* constructs that had inhibited MSV replication in BMS, and transgenic lines were tested for virus resistance. Third, the maize cultivar Hi-II was transformed with the constructs that resulted in MSV-resistant *D. sanguinalis*.

2.2 MATERIALS AND METHODS

2.2.1 Clone Construction[ξ]

All *Rep*-based clones were derived from an infectious MSV plasmid, pKom602, which is a partial tandem dimer of the MSV-Kom genome cloned in pUC19 (Schnippenkoetter *et al.*, 2001). The C1/C2 ORFs of pKom602 were amplified by PCR using the forward primer C1: 5' TTAGGATCCCTCAGCCTCAACCCTCC, which introduced a *Bam*HI restriction enzyme site (underlined) 26 bp upstream of the C1 start codon, and the reverse primer C2: 5' ACGCAAACAATACAGGGGGGTAGTC, which binds in the SIR. The PCR product was cloned into the *Bam*HI/ *Hind*II site of pBluescript SK+ (pSK; Stratagene, La Jolla, CA) and all subsequent mutations and truncations were performed on this wild type (wt) Rep construct (pSK*Rep*). Unless otherwise stated, all restriction enzymes were obtained from Boehringer Mannheim (Mannheim, Germany). Sequencing of clones was carried out by D. James at the University of Cape Town (UCT), using an ALF Express automated sequencer (Pharmacia Corporation, Peapack, NJ). Sequence analysis was carried out using DNAMAN (version 4.0; Lynnon BioSoft, Quebec). Standard cloning techniques were used as in Sambrook *et al.* (1989).

2.2.1.1 PCR site-directed mutagenesis

As part of a detailed analysis of the effects of changing amino acids within various Rep motifs, pSK*Rep* was used as template DNA for mutagenic PCR (T. Mangwende, 2001). In the present study, Rep constructs containing novel mutations in motif III and the RBR protein interaction domain of MSV Rep were chosen for transformation of plants because they completely abolished MSV replication in black Mexican sweetcorn (BMS). Mutagenic primer sequences and introduced enzyme sites used to detect the mutations are shown in Table 2.1. All mutated clones were confirmed to be correct by sequence analysis.

TABLE 2.1 Mutations introduced into pSK*Rep*.

Mutagenic Primer Sequence (altered nucleotides are shown in lowercase)	**Amino Acid Changes**	**Introduced enzyme site**	**Mutated clone name**
5'-GATTTACTTTGTctTaAGTCAATCAAC-3'	[a]LLCN^{201}E^{202} to LLCL^{201}K^{202}	*Bfr*I	pSK*Rep*$^{III+Rb-NTP+}$
5'-AGAGTGAGGGccTAttTTCTCAAGGAAC-3'	[b]VRD^{99}YI^{101}LKE to VRA^{99}YF^{101}LKE	*Hae*III	pSK*Rep*$^{III-Rb+NTP+}$

[a]Mutation introduced into the Rep RBR-interaction domain. Amino acid numbering is relative to the Rep start codon.
[b]Mutation introduced into the Rep motif III. Amino acid numbering is relative to the Rep start codon.

[ξ] Properties of all plasmids and constructs used in Chapters 2 to 4 are summarized in Appendix A.

2.2.1.2 Truncated Rep constructs

To generate C-terminal truncated Rep genes, pSK*Rep* (wt) and pSK*Rep*$^{III\text{-}Rb+NTP+}$ (III$^-$ mutant) were digested with *Bam*H1/*Hind*III, creating 537-bp N-terminal fragments that were subsequently cloned into the *Bam*HI/*Hind*III site of pSK. The resulting plasmids were called pSK*Rep*$^{\Delta Rb\Delta C2}$ and pSK*Rep*$^{III\text{-}\Delta Rb\Delta C2}$ respectively (see Fig.2.1 for a diagrammatic illustration). In addition, pSK*Rep* and the Rb$^-$ mutant, pSK*Rep*$^{III+Rb\text{-}NTP+}$, were subjected to a partial digest with *Bam*H1/*Hind*III, and the resulting 658 bp N-terminal fragments cloned into pSK to create the plasmids designated pSK*Rep*$^{\Delta C2}$ and pSK*Rep*$^{Rb\text{-}\Delta C2}$ respectively (Fig. 2.1). All constructs were confirmed to be correct by sequencing.

2.2.1.3 Construction of plant vectors for *trans*-replication analysis of mutant and truncated Rep proteins

A 1.3-Kb *Bam*HI-*Bgl*II fragment (containing the full-length Rep gene; see Fig 2.1) from each of the plasmids pSK*Rep*$^{III\text{-}Rb+NTP+}$ (III$^-$ mutant), pSK*Rep*$^{III+Rb\text{-}NTP+}$ (Rb$^-$ mutant), pSK*Rep*$^{III\text{-}Rb\text{-}NTP+}$ (III$^-$ Rb$^-$ double mutant) (T. Mangwende, 2001) and pSK*Rep* (wt Rep gene; this thesis) was cloned into the *Bam*HI site of a 5' dephosphorylated plasmid, pAHC17 (Christensen and Quail, 1996). The resulting plasmids were designated p*Rep*^{Mut}, where Mut = the III$^-$, Rb$^-$, and III$^-$Rb$^-$ mutations (T. Mangwende, 2001), or p*Rep* (wt *Rep*). To create an antisense Rep gene, pSK*Rep* was cut with *Bam*HI and *Bgl*II to release the full-length Rep gene, and cloned into the *Bam*HI site of pAHC17 in the antisense orientation (designated p*Rep*$^{III+Rb+NTP+(AS)}$; T. Mangwende, 2001). To clone the truncated Rep genes into pAHC17 (this thesis), a *BamH*I site was inserted at the C-terminus of the truncated genes. The truncated genes were then cut with *Bam*HI, and inserted into the same site of pAHC17. Vectors with the Rep genes in the sense orientation were selected, and designated p*Rep*$^{\Delta}$ or p*Rep*$^{Mut\Delta}$, where Δ= a deletion, and Mut = Rb$^-$ or III$^-$ mutations. To assay the effect of truncated antisense Rep genes on viral replication, *Rep*$^{\Delta C2}$ and *Rep*$^{\Delta Rb\Delta C2}$ cloned in the *Bam*HI site of pAHC17 in the antisense orientation were selected and designated p*Rep*$^{\Delta C2(AS)}$ and p*Rep*$^{\Delta Rb\Delta C2(AS)}$ respectively.

In most cases, transformation of *D. sanguinalis* was carried out by co-bombardment of p*Rep*^{Mut} or p*Rep*$^{III+Rb+NTP+(AS)}$ or p*Rep*$^{Mut\Delta}$ (all in pAHC17) with pAHC25, which contains the bar (bialophos resistance) gene and the GUS (*uidA*) gene under the control of separate maize ubiquitin promoters (Ubi-Bar/ Ubi-Gus; Christensen and Quail, 1996). However, with the co-bombardment strategy, both plasmids have to integrate into the genome of one cell to regenerate bialophos resistant/ *Rep* transgenic plants. To dispense with co-bombardment, the GUS gene in

pAHC25 was replaced with $Rep^{III\text{-}Rb\text{-}NTP+}$. The Rep gene was isolated from $pRep^{III\text{-}Rb\text{-}NTP+}$ as a *Pst*I (whose sites flank the Rep gene in pAHC17) fragment, and the GUS gene was excised out of pAHC25 with *Sma*I and *Sac*I. The $Rep^{III\text{-}Rb\text{-}NTP+}$ fragment was then blunt-cloned into the *Sma*I/*Sac*I site of pAHC25. To do this, the *Sma*I/*Sac*I-digested vector and the *Pst*I-Rep fragment were first treated with Klenow DNA polymerase, followed by dephosphorylation of the vector with SAP (shrimp alkaline phosphatase). The $Rep^{III\text{-}Rb\text{-}NTP+}$ fragment was then ligated into the dephosphorylated/ blunt ended vector. $Rep^{III\text{-}Rb\text{-}NTP+}$ in the sense orientation was selected, and designated $pAHCRep^{III\text{-}Rb\text{-}NTP+}$ (Ubi-Bar/ Ubi-Rep).

2.2.1.4 Construction of intronless Rep and RepA

pSK*Rep* was used as a template to create an intronless Rep gene, by inverse PCR. The forward primer RepΔI-F (5' -TCATCAGATGAAAGATCAAGAAAGC - 3') amplified the Rep gene from the 3' end of the intron through the C2, while the reverse primer RepΔI-R (5'-CTGGAAGATGTTAGGCTGGAGCC - 3') amplified the gene from the 5' end of the intron through the C1. In this way the whole template plasmid was amplified, minus the intron. The PCR product was self-ligated to create the intronless plasmid $pSKRep^{\Delta I}$, which was then confirmed to be correct by sequencing. The same procedure was followed to create an intronless $RepRb^-$ gene, this time using $pSKRep^{Rb-}$ as the template, creating the plasmid $pSKRep^{Rb\text{-}\Delta I}$.

The RepA gene contains the entire intron at its C-terminus, therefore the first step in making a *RepA* construct was to prevent the possibility of splicing occurring. This was done using a *Rep* plasmid (pMB1657, provided by Dr. M. Boulton, John Innes Centre, Norwich, U.K.) with a 3' splice site mutation of $A^{733}G^{734}$ to $T^{733}C^{734}$ (Wright *et al.*, 1997). A fragment containing the splice site mutation was excised from pMB1657 with *Xho*I and *Bgl*II, and swapped with the same wt fragment from pKom602, resulting in pKomMB1657. Although pMB1657 contains the MSV-Ns strain, the *Xho*I/*Bgl*II fragment is 100% homologous to the same fragment from MSV-Kom. Therefore, the only mutation introduced into pKom602 by the fragment swap was the desired Rep 3' splice site mutation. The RepA gene was amplified from pKomMB1657 using the C1 forward primer, which introduced a *Bam*H1 site at the 5' end, and a reverse primer (*RepABgl*, provided by D. McGivern, John Innes Centre, Norwich, U.K.), which introduced a *Bgl*II site immediately after the *RepA* stop codon. The *RepA*Bgl primer has the following sequence: 5' - TTATAGATCTCTAGGCTTCTGG - 3'. Once amplified, the RepA gene was cloned into the *Bam*H1 site of pSK to form pSK*RepA*, and subsequently confirmed to be correct by sequencing.

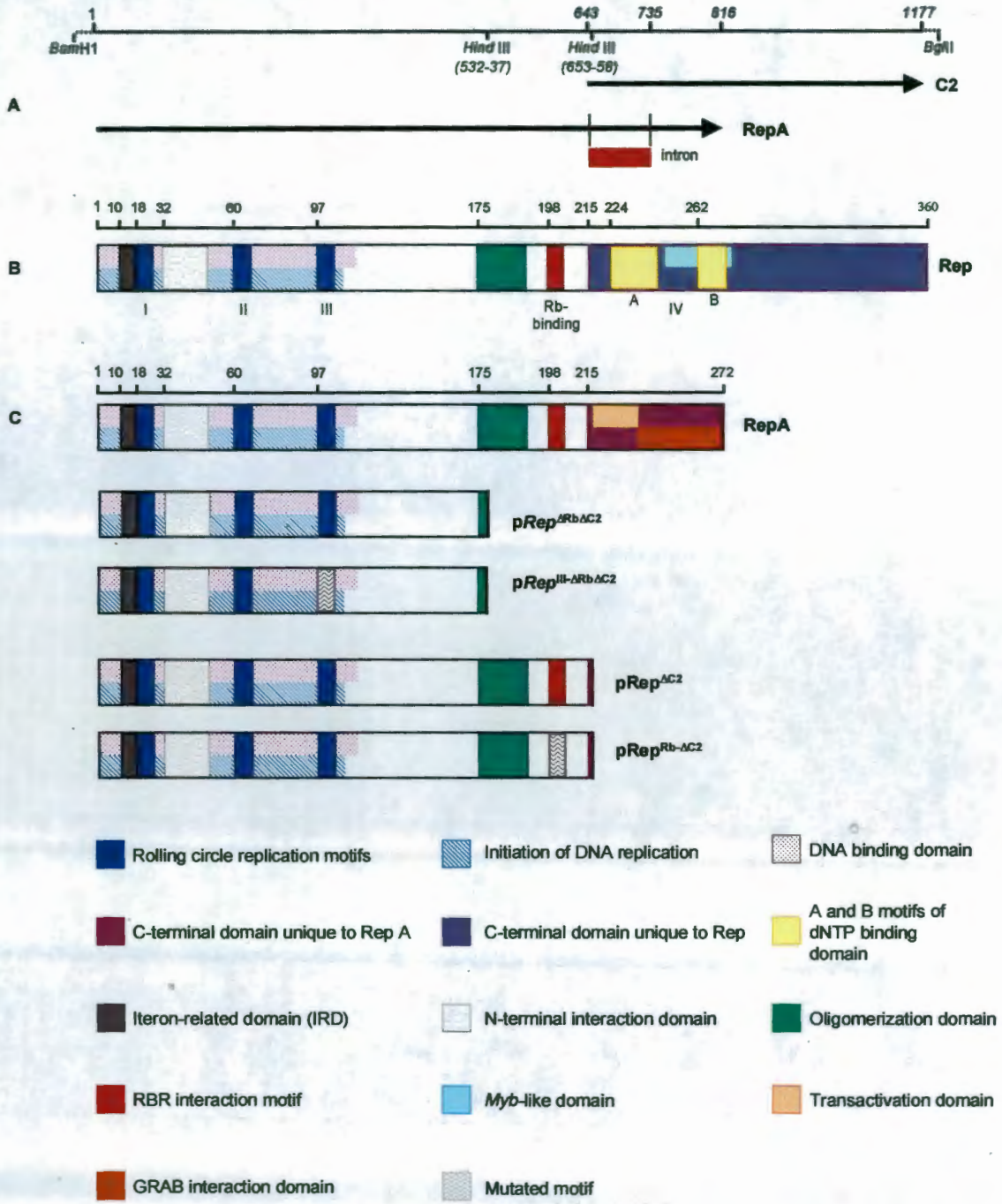

FIGURE 2.1 MSV Rep genes and gene products expressed in transient replication assays. (A) The MSV-Kom RepA (C1) and C2 ORFs are shown, including selected important nucleotide co-ordinates. Numbering of nucleotides is relative to the Rep ATG (A = 1). The positions of restriction sites used in the cloning of various Rep genes are also shown (see text for details). The *Bam*H1 site was introduced upstream of the *Rep* ATG, and the *Bgl*II site is in the MSV SIR. (B) The known sequence motifs and functional domains of the mastrevirus Rep protein (expressed from the spliced C1/C2 ORFs). Amino acid numbering is relative to the Rep start codon. (C) Truncations of RepA (shown relative to the full-length RepA protein) used in the replication assays. The motifs present in each Rep variant are shown, with mutated motifs represented by grey zigzags.

The same procedure was followed to obtain the RepA version of the Rb$^-$ mutant, using pKomRb$^-$

(described in chapter 3) as a starting template in place of pKom602. The resulting RepA plasmids, both containing the 3' splice site mutation, were called pSK*RepA* (wt) and pSK*RepA*$^{Rb-}$ (Rb$^-$ mutant)

2.2.1.5 Clone construction for yeast two-hybrid analysis

The Rep genes of pSKRep$^{\Delta I}$, pSKRep$^{Rb-\Delta I}$, pSKRepA, pSKRepA^{Rb-}, pSKRep$^{\Delta C2}$ and pSKRep$^{Rb-\Delta C2}$ were cloned in frame with the GAL4 activation domain into pGAD424 (CLONTECH, CA) and in frame with the GAL4 binding domain into pBD-GAL4-CAM (Stratagene, La Jolla, CA) to create Rep-GAL4 fusion products. To create the appropriate enzyme sites in order to clone the Rep genes in frame with the GAL4 binding domain, primers were used (Table 2.2A) which added on the *Sal*I restriction site (forward primer) and the *Pst*I restriction site (reverse primer) to the 5' and 3' ends respectively of each Rep gene. The same procedure was followed to create the appropriate enzyme sites in order to clone the Rep genes in frame with the GAL4 activation domain (Table 2.2B), this time adding on *Sal*I to the 5' end and *Bgl*II to the 3' end of each Rep gene. Rep genes in pGAD424 were designated pAD*Rep*$^{\Delta}$, pAD*Rep*$^{Mut\Delta}$ or pAD*Rep*^{Mut}, while those cloned in pBD-GAL4-CAM were designated pBD*Rep*$^{\Delta}$, pBD*Rep*$^{Mut\Delta}$ or pBD*Rep*^{Mut} (where Δ= a deletion, and Mut = Rb$^-$ mutation). All constructs were confirmed to be correct by sequencing.

TABLE 2.2 Primers used in the cloning of various Rep genes into (A) pBD-GAL4-CAM and (B) pGAD424

(A) pBD-GAL4-CAM	**Forward Primer for all Rep genes:**
	BD*Sal*C1(F) 5'-CCGGGTCGACTCATGGCCTCCTCCTCATCC-3'
***Rep* Template**	**Reverse Primer**
pSKRep$^{\Delta I}$, pSKRepRb$^{-\Delta I}$	BD*Pst*C2 5'-CTTGGCTGCAGTTTACACTTCCTCCGTAGGAGG-3'
pSKRep$^{\Delta C2}$, pSKRep$^{Rb-\Delta C2}$	BD*Pst*ΔC2 5'-CTTGGCTGCAGTAAGCTTCGGGACTAACCTGG-3'
pSKRepA, pSKRepA^{Rb-}	BD*Pst*RepA 5'-CTTGGCTGCAGTCTAGGCTTCTGGCCCAAG-3'
(B) pGAD424	
***Rep* Template**	**Reverse Primer**
pSKRep$^{\Delta I}$, pSKRep$^{Rb-\Delta I}$	AD*Bgl*C2 5'-TTCATAGATCTCTTACACTTCCTCCGTAGGAGG-3'
pSKRep$^{\Delta C2}$, pSKRep$^{Rb-\Delta C2}$	AD*Bgl*ΔC2 5-TTCATAGATCTCAAGCTTCGGGACTAACCTGG-3'
pSKRepA, pSKRepA^{Rb-}	AD*Bgl*RepA 5'-TTCATAGATCTCCTAGGCTTCTGGCCCAAG-3'

The cDNA of the maize RBR protein (pZmRb1; Xie *et al.*, 1996) cloned into pGBT9 (pGBT9ZmRb1) was provided by Dr G. Horvath (described in Horvath *et al.*, 1998).

2.2.2 Yeast Two-Hybrid Analysis of the RBR Protein Interaction Properties of Mutant, Truncated and Wild Type *Rep* Constructs

While the cloning of the *Rep* constructs described above for yeast two-hybrid analysis was done as part of this thesis, the actual yeast transformation and analysis of the RBR-interaction properties of the *Rep* constructs was performed by D. McGivern (John Innes Centre, Norwich, U.K.) as described below.

Plasmids containing *GAL4* binding domain fusions (*trp1* transformation marker) and *GAL4* activation domain fusions (*leu2* transformation marker) were introduced separately into *Saccharomyces cerevisiae* strains CG1945 and Y187 respectively as described by Gietz and Woods (1994). The transformation mixture was plated onto yeast drop-out selection media lacking the appropriate amino acid to select for transformants. Yeast strains CG1945 (MATa; transformed with pZmRb1) and Y187 (MATα; transformed with Rep gene derivatives) were mated according to a protocol modified from the CLONTECH Yeast Protocols Handbook (CLONTECH Laboratories Inc., Palo Alto, Ca, USA). The mated yeast was grown on selective drop-out medium lacking tryptophan (Trp) and leucine (Leu), and drop-out medium lacking Trp, Leu and histidine (His) and containing 5 mM 3-amino-1,2,4-triazole (3AT). Only strains containing interacting fusion proteins can grow on the latter medium.

2.2.3 Transient Replication Assays

2.2.3.1 Maintenance of black Mexican sweetcorn (BMS) suspension cultures

BMS suspension culture cells were maintained in the dark at 26°C with constant rotary shaking (120 r.p.m.) in BMS liquid medium (pH 5.75), which is MS medium (Murashige and Skoog, 1962) supplemented with 3% sucrose, 0.01% myoinositol and 2 mg/L of 2,4-dichloro-phenoxyacetic acid (2,4-D). Cells were subcultured at a 1:6 dilution every two weeks.

2.2.3.2 Transfection of BMS by particle bombardment

BMS cells were subcultured at a 1:3 dilution three days prior to bombardment. Four hours before bombardment, 1.0 mL packed volume of cells was filtered onto Whatman # 4 filter paper, which

was placed on BMS high osmoticum solid medium (BMS medium containing 0.2 M mannitol, 10 μg/mL silver nitrate and 0.8% agar). An aliquot of 1 μg of each plasmid was precipitated onto 1 μm gold particles (50 μL of 60 mg/mL gold suspended in 50% glycerol) according to the protocol of Dunder *et al.* (1995). Cells were bombarded using a Biorad/DuPont PDS1000-He system at a pressure of 650 psi under a vacuum of 27 inHg. The gap distance was 6 mm, the macrocarrier travel distance 5 mm, and the target distance 6 cm. Cells on each plate were bombarded twice, each shot delivering approximately 167 ng of each plasmid. In each replication experiment, nine plates of BMS were co-bombarded with pKom602 and one of the Rep constructs, p*Rep*^{Wt}, p*Rep*^{Mut}, p*Rep*$^{\Delta}$ or p*Rep*$^{Mut\Delta}$ provided in *trans*, and nine plates of BMS were co-bombarded with pKom602 and pAHC17. Since pAHC17 is the vector in which the Rep genes were cloned, it provided a non-Rep co-bombardment control. From here on, co-bombardment of pKom602 and pAHC17, with no Rep gene provided in *trans*, is referred to as bombardment of "pKom602 alone". After bombardment, cells were incubated in the dark at 26°C for 24 hours, after which they were transferred to BMS solid medium and incubated in the same conditions for four days.

2.2.3.3 Quantitative PCR

Total DNA was extracted from BMS cells four days after bombardment, using the method of Dellaporta *et al.* (1983) except that, having resuspended the chromosomal DNA in 50 mM Tris-HCl, 10 mM EDTA (pH 8), 600 μg/mL RNaseA was added and the mixture incubated for 1 hour at 37°C. After a second precipitation with isopropanol, DNA was finally resuspended in water and diluted to 50 ng/μL. To ensure all samples were the same concentration, equal amounts of total DNA were electrophoresed through a 0.8% agarose gel, and band intensity measured using the virtual densitometer computer program, GelTrak (Dennis Maeder, University of Cape Town). The relative amounts of viral replication were determined using a quantitative PCR-based assay (Fig. 2.2). Primers were designed to amplify the region corresponding to nucleotides 1595-209 in the MSV-Kom genome (relative to the unique *Bam*H1 site at the beginning of the V1 gene). Primer sequences were: MSV-DEG1 5'-TTGGVCCGMVGATGTASAG -3' and MSV-DEG2 5'-CCAAKDTCAGCTCCTCCG -3' (Willment *et al.*, 2001). These primers, which are able to amplify viral DNA once it has replicated, cannot amplify linear MSV-Kom DNA from the pKom602 input plasmid (see Fig 2.2). To confirm this, total DNA extracted immediately after bombardment (as a control for input plasmid DNA amplification) was subjected to the same PCR. Each PCR reaction was "spiked" with pKep177 (obtained from K.E. Palmer, UCT) of known concentration as an internal control. The amount of pKep177 added to each reaction was

in the range of 20 to 100 pg, and the amount of total DNA was 100ng, depending on the level of viral replication approximated previously by non-quantitative PCR. Optimum ratios of spike to viral DNA were required to avoid the one out-competing the other. The relative concentration of replicated DNA was calculated by determining the ratio of the replicative form (RF) virus band intensity to the pKep177 competitor band intensity, using GelTrak. pKep177 has a *Pst*I site inserted 72 bp downstream from the start codon, which distinguishes it from the RF viral band. The principle of the quantitative PCR assay is explained in Fig. 2.2.

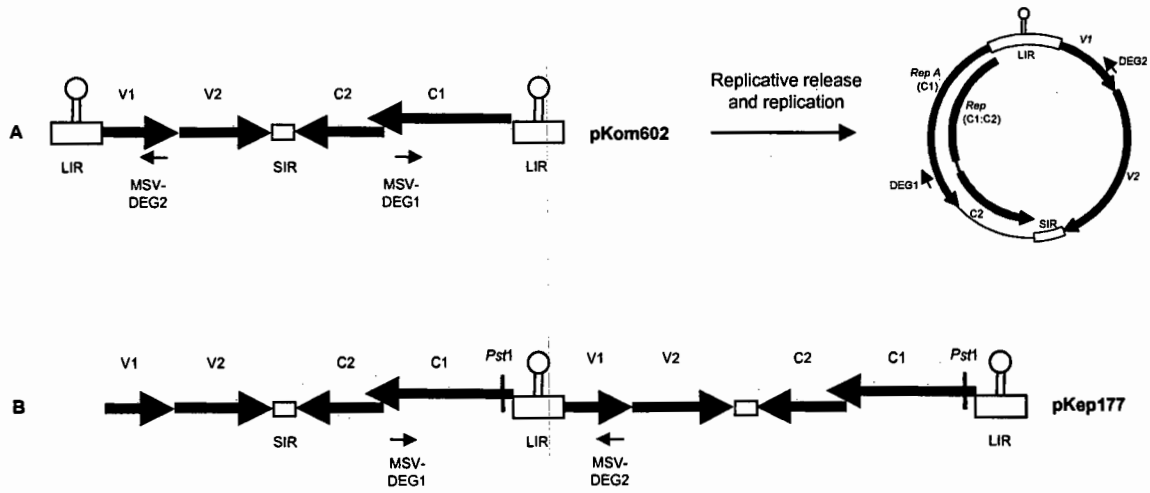

Figure 2.2 Plasmids and primers used for quantitative PCR assay. (A) The primers MSV-DEG1 and MSV-DEG2 do not amplify linear MSV-Kom DNA from pKom602. In the nuclei of BMS cells the viral DNA is replicatively released from the vector, and as circular RF DNA it can serve as a template for the amplification of a 1314 bp fragment. (B) Included in the PCR reaction is an internal control, pKep177, of known concentration. This competitor is a tandem dimer of the MSV-Kom genome, with a *Pst*I site inserted at the start of the *C1* ORF. Being a dimer, pKep177 is amplified by the MSV-DEG1 and MSV-DEG2 primers and competes with viral RF DNA for primers and other PCR components. Digestion of amplified pKep177 with *Pst*I yields two bands of 604 and 710 bp, allowing the competitor to be distinguished from amplified MSV-Kom DNA. The relative concentration of replicated viral DNA is calculated by determining the ratio of MSV-Kom DNA band intensity to that of pKep177, whose concentration is known.

2.2.4 *Digitaria sanguinalis* Tissue Culture

2.2.4.1 Callus induction

Young unemerged inflorescences of *D. sanguinalis* were surface-sterilised by soaking in 70% ethanol for 1 minute followed by 0.35% sodium hypochlorite for 20 minutes, and were then washed four times with sterile distilled water. The inflorescences were cut into 5 mm segments and transferred to callus induction medium, which is MS medium (adjusted to pH 5.8) supplemented with 3% sucrose and 2.5 mg/ L 2,4-D, and solidified with 0.8% agar. Plates

containing the inflorescences were maintained in the dark at 26°C, as were the calli once they had initiated.

2.2.4.2 Transformation of *D. sanguinalis*

Embryogenic *D. sanguinalis* calli were transformed by particle bombardment using the Biorad/ DuPont PDS1000-He system. Sixteen hours prior to bombardment, embryogenic calli were transferred to high osmoticum medium (MS medium containing 2.5 mg/L 2,4-D, 100 mg/L myo-inositol, 0.2 M mannitol and 10 mg/L silver nitrate). For each bombardment, two μg of plasmid DNA were precipitated onto gold particles as described for transfection of BMS. Each mutant or truncated *Rep* construct was co-bombarded with pAHC25 (Ubi-Bar/UbiGus) at a 1:1 weight ratio. *Rep* plasmids chosen along with pAHC25 to co-transform *D. sanguinalis* were p*Rep*$^{III\text{-}Rb+NTP+}$, p*Rep*$^{III\text{-}Rb\text{-}NTP+}$, p*Rep*$^{Rb\text{-}\Delta C2}$ and p*Rep*$^{III+Rb+NTP+(AS)}$. However, in bombardments with pAHC*Rep*$^{III\text{-}Rb\text{-}NTP+}$ (Ubi-Bar/Ubi-Rep), there was no co-bombardment with pAHC25. The settings on the biolistics device were as follows: the gap distance was 6 mm, the macrocarrier travel distance 5 mm, and the target distance 6 cm. Each target plate was bombarded twice at a pressure of 900 psi, each shot delivering approximately 333 ng per plasmid. Twenty-four hours after bombardment the calli were transferred from high osmoticum to MS maintenance medium. Non-bombarded calli, and calli bombarded with pAHC25 alone were used as controls in all experiments.

2.2.4.3 Selection and regeneration of transgenic calli

Seven days after bombardment, calli were transferred to regeneration medium (MS medium containing 0.1 mg/L napthaleneacetic acid and 10 mg/L benzylaminopurine [NAA and BAP, Sigma]) with selection (3 mg/L bialophos). Once on regeneration/ selection medium, calli were kept in the dark for five days at 26°C, followed by 16 hours diffuse light for nine days, after which they were exposed to 16 hours full light per day. Shooting callus was then transferred to rooting medium (MS medium without hormones or plant growth regulators). Once roots had grown, plantlets were hardened off in a 33:33:33 mix of sand, compost and palm peat and finally transferred to potting soil.

2.2.4.4 Seed germination

Seed collected from transgenic plants were incubated at 37°C overnight. They were then surface sterilised by shaking in 70% ethanol for five minutes, followed by 3.5% sodium hypochlorite, 0.02% Triton X-100 for ten min, and then washed five times with sterile distilled water. The

sterilised seeds were suspended in 5 mL of 0.1% agar and pipetted onto a plate containing PNS medium (Plant Nutrient agar medium with Sucrose (pH 5.5) solidified with 0.75 % nutrient agar. The seeds on PNS plates were kept in the light at 26°C until they germinated. The plantlets were then hardened off as described for *D. sanguinalis* regeneration.

2.2.5 Transgene Expression Analysis

2.2.5.1 Reverse transcription-PCR (RT-PCR)

Total RNA was extracted from stock transgenic callus material (previously initiated from T_0 plants of each transgenic line), using the reagent TRIzol® (GIBCOBRL). ~200 mg of callus material was homogenized in 1 mL Trizol, followed by incubation at room temperature (RT) for 5 min. Insoluble material was removed from the homogenate by centrifugation at 12,000 g for 10 min at 4°C. 200 µL of chloroform was added to the supernatant and shaken vigorously for 3 min at RT. The samples were centrifuged at 12,000 g for 15 min at 4°C. RNA was precipitated by the addition of 0.5 mL isopropanol to the supernatant, which was incubated at RT for 10 min, followed by centrifugation at 12,000 g at 4°C. The RNA pellet was washed with 1 mL 75% ethanol, and resuspended in 50µl DEPC-treated water (that is 0.5 mL diethyl pyrocarbonate [Sigma] added to 1 L distilled water, left to stand overnight, and autoclaved).

As a positive control for the RT-PCR reaction, the Rep gene (from pGEMT*Rep*, cloned in frame with the T7 promoter in the pGEM®-T Easy Vector (Promega, WI) was transcribed using the Ribomax™ RNA production system (Promega, WI). pGEMT*Rep* was linearized with *Pst*1, precipitated from the digest reaction mixture using 4M LiCl, and the resuspended DNA was used in the transcription reaction, as described in the kit manual (also available at www.promega.com). After transcription, 1 µL of DNAse (Promega, WI) was added to the transcription reaction and incubated for 30 min at 37°C to remove the original DNA template, after which the DNAse was inactivated at 90°C for 5 min.

The RNA from each transgenic sample was also DNAse-treated as described above before undergoing RT-PCR, to remove any contaminating transgene DNA that could be amplified by the DNA polymerase. A control reaction was carried out using the same RT-PCR conditions except that the reverse transcriptase enzyme was omitted to confirm that amplification was not due to contaminating genomic DNA. Primers were designed that could amplify all Rep transgenes (truncated and full-length) from the different transgenic samples. These were RepΔC2F: 5' ATGGCCTCCTCCTCATCCAAC 3' and RepΔC2R: 5' AAGCTTCGGGACTAACCT 3'. The

bar transcript was amplified from each transgenic sample by RT-PCR using the primers: BAR1 (forward) 5' CGTCAACCACTACATCGAG 3' and BAR2 (reverse) 5' GAAACCCACGTCATGCCAG 3' obtained from T. Dube, UCT.

RT-PCR was carried out using the Access RT-PCR System (Promega, WI). The RT-PCR cycles were as follows: 48°C for 30 min (first strand synthesis); 94°C for 2 min (inactivation of reverse transcriptase and denaturation of primers/cDNA); [94°C for 30 sec (denaturation), 60°C for 30 sec (annealing), 68°C for 1 min (extension)] x40 cycles; 68°C for 2 min (final extension).

2.2.5.2 GUS assays

Expression of the GUS gene (*uidA*, encoding β-glucuronidase) in both transient and transgenic assays was analysed by histochemical and protein (spectrofluorometric) assays. In the histochemical assay, transient GUS activity in *D. sanguinalis* calli that had been bombarded with pAHC25 was visualized by addition of the GUS substrate X-gluc (Sigma) to the calli three days after bombardment (described in Jefferson *et al.*, 1987). GUS expression in transgenic calli and in leaves from transgenic plants was analysed in the same way. In the spectrofluorometric assay, protein was extracted from calli three days after bombardment, and GUS activity was determined according to Jefferson (1987). Protein concentrations were determined by the method of Bradford (1976) and GUS activity was corrected for protein concentration.

2.2.6 Test for Resistant *D. sanguinalis* by Challenge with MSV

Transgenic plants were challenged with MSV using viruliferous leafhoppers (*C. mbila*), obtained from Dr. Mike Barrow (Pannar Ltd, Greytown, South Africa). The leafhopper population at Pannar Ltd had been fed on symptomatic maize leaves collected from various locations in South Africa and Zimbabwe in order to develop a mixed (albeit unknown) MSV population, and were subsequently maintained at UCT on *Zea mays* cv Jubilee. In the challenge experiments, the viruliferous leafhoppers were placed in a small vial containing a ~1 mm slit, through which a single leaf from a transgenic plant was inserted, ensuring that the leafhoppers in a particular vial could only feed on one leaf of one plant. Four leafhoppers were placed in each vial, and three vials were placed at different positions on each plant. Thus, there were three points of entry for the virus, and in total 12 leafhoppers feeding on each plant. One challenge experiment, using plants transgenic for p*Rep*$^{\text{III-Rb-NTP+}}$, was carried out at Pannar Ltd in Greytown, South Africa, under the supervision of Dr Rikus Kloppers. In that case, five plants from each of two lines and a non-transgenic control plant were challenged by feeding leafhoppers on three different leaves of

each plant, as explained above. The rest of the challenge experiments, carried out at UCT, Cape Town, were slightly different. Leafhoppers in vials were fed first on a non-transgenic or pAHC25-transformed plant for two days, then transferred to a transgenic plant and allowed to feed for two days. Since the same leafhoppers feed on both the transgenic and control plants, in the case of a transgenic plant not becoming infected while the control does get infected, this method ensures that the lack of infection is not due to non-viruliferous leafhoppers, but rather to the plant being resistant to viral infection.

2.2.7 Maize (Hi-II) Tissue Culture

Embryogenic high type II (Hi-II) calli, obtained from Dr. W. Gordon-Kamm (Pioneer Hi-Bred, International, Inc., Johnston) were maintained on N6 medium (Chu, 1978), pH 5.8, supplemented with 3% sucrose, 0.3% proline, 0.01% casamino acids and 2 mg/L 2,4-D, solidified with 2.4g/L Gelrite.

2.2.7.1 Transformation of Hi-II

Actively growing embryogenic calli were selected as target tissue for gene delivery by particle bombardment, using the Biorad/ DuPont PDS1000-He system. The calli were placed onto target plates containing high osmoticum medium (N6 maintenance medium with 36.4 g/L mannitol, 36.4 g/L sorbitol and 10 mg/L silver nitrate) four hours prior to bombardment. Two μg of plasmid DNA were precipitated onto gold particles as described for transfection of BMS. The *Rep* plasmid chosen to transform Hi-II, p*Rep*$^{Rb\text{-}\Delta C2}$, was co-bombarded with pAHC25 (Christenson and Quail, 1996) at a pressure of 1100 psi. The biolistic device settings were as follows: 8 mm between the rupture disc and macrocarrier, 10 mm between the macrocarrier and the stopping screen, and 7 mm between the stopping screen and the target. After the first shot, a recovery time of four hours was allowed before the calli were shot a second time. Approximately 333 ng of each plasmid were delivered per shot. Twenty-four hours after bombardment the calli were transferred from high osmoticum to maintenance N6 medium. Non-bombarded calli, and calli bombarded with pAHC25 alone were used as controls in all experiments.

2.2.7.2 Selection and regeneration of transgenic calli

Four days after bombardment, calli were transferred to a gentle selection medium (N6 maintenance medium with 1 mg/L bialophos) on which they were maintained for two weeks. All subsequent selection was more stringent at 3 mg/L bialophos. Selection continued for 6-8 weeks

in the dark, after which bialophos-resistant calli were transferred to shoot initiation medium (MS medium with 5 mg/L BAP, 0.25 mg/L 2,4-D, 3% sucrose and 3 mg/L bialophos). After five days in the dark followed by nine days under diffuse light, the calli were kept in the light for 16 hours a day. Once shoots emerged they were transferred to shoot elongation medium (shoot initiation medium without BAP or 2,4-D) and finally root elongation medium (MS supplemented with 1.0% sucrose and 3 mg/L bialophos). Stringent selection was maintained during the whole regeneration process. Once roots had grown, plantlets were hardened off in a 33:33:33 mix of sand, compost and palm peat and finally transferred to potting soil. Adult transgenic plants were sent to Pannar Ltd, Greytown, SA, for controlled self-pollination and setting of seed.

2.2.8 Analysis of the MSV Strain Composition Transmitted by Viruliferous Leafhoppers.

0.5 g to 2 g of infected maize leaf material was frozen in liquid nitrogen and finely ground. 6 mL of Extraction Buffer (0.1 M Tris-HCl; 0.1 M NaCl; 0.1 M EDTA; 1% SDS w/v; pH7) was added, the mixture shaken and incubated at 65 °C for 5 min. After centrifuging at 10 000 G for 5 min, 5.5 mL of the supernatant was added to 5.5 g CsCl and shaken until the CsCl dissolved, followed by centrifugation at 10 000G for 5 min at RT. 5 mL of the supernatant, plus 10 µg of pSK and 400 µl of 10 mg/mL ethidium bromide were added to a 5 mL Quick-Seal tube and centrifuged in a Beckman Vti65 rotor at 266 805 G overnight at 20°C. After the separation of linear ds DNA (plant genomic DNA, upper band) from ds ccDNA (viral and plasmid DNA, lower band) the lower band, visualized using ultra violet (UV) light at 310 nm, was collected from the tube using a sterile wide-bore needle on a sterile 2 mL syringe. The ethidium bromide was extracted from the collected fraction by the addition of an equal volume of salt-saturated phenol (300 mL of 5M NaCl in TE buffer [10mM Tris, 1mM EDTA] added to 600 mL of isopropanol), followed by thorough mixing and centrifugation for 1 min in a microcentrifuge. The upper isopropanol phase containing the ethidium bromide was discarded. This procedure was repeated twice, before adding two volumes of sterile distilled water to the DNA-containing phase. The DNA was pelleted by the addition of 1 volume isopropanol, incubation on ice for 10 min and centrifugation for 15 min. The pellet was washed in 70% ethanol and resuspended in sterile distilled water.

50 µL of DNA (of unknown concentration) was digested with *Bam*HI, followed by precipitation of the DNA with 4 M LiCl. The DNA was then cloned into the *Bam*HI site of pSK and transformed into *E. coli* DH5α. Colonies were picked and replica-plated onto two plates, one containing solidified bacterial growth medium with ampicillin selection and the other containing

a Hybond-N+ membrane (Amersham Pharmacia, UK). 100 colonies were screened by colony hybridization using DIG-labelled (Boehringer Mannheim) whole genome probes. An equal mixture of four probes consisting of the genomes of MSV-Kom, MSV-Set, MSV-Mat and MSV-VW was used. The colony hybridization was carried out according to the Hybond-N+ manufacturer's instructions (Amersham Pharmacia). Chemiluminescent detection of colonies containing DNA that hybridized to the genome probes was carried out according to the DIG System User's Guide (Boehringer Mannheim). Positive colonies picked from the replica-plated bacteria on growth medium were subjected to PCR using primers that annealed to either side of the pSK polycloning site. Primer sequences were: M13 uni Rev 5' AGCGGATAACAATTTCACACAGG 3' and M13 uni Fwd 5' CCCAGTCACGACGTTGTAAAACG 3'. PCR-amplified fragments were analysed by restriction length polymorphism (RFLP) to distinguish different virus strains from one another, as described in Willment *et al.* (2001). Plasmid DNA was extracted from colonies containing different MSV genomes identified by RFLP of the PCR products, and the whole genomes were sequenced, first using the M13 Fwd and Rev primers described above, and then internal primers in the forward and reverse directions. Using DNAMAN, the sequences were aligned with a variety of African streak virus genomic sequences, and the relationship between the genomes was determined using a rooted neighbour joining tree.

2.3 RESULTS AND DISCUSSION

2.3.1 Interactions between MSV Rep and RBR proteins

Since some of the *Rep* constructs used in the replication assays included mutations in the RBR protein interaction domain, it was important to first determine whether the RBR protein-binding ability of these mutants was abolished. In order to separate Rep from RepA, intronless *Rep* (in effect behaving as a spliced *Rep* and therefore capable of expressing Rep only) and *RepA* genes were made, both wt and with the Rb$^-$ mutation. The ability of truncated, mutated and wt Rep and RepA gene products to bind to the maize RBR protein (ZmRb1) was determined using a yeast two-hybrid assay (Table 2.3). Interaction of a Rep product fused in-frame to the Gal4 activation domain with ZmRb1 fused in-frame to the GAL4 binding domain allows growth of yeast on drop-out media lacking histidine, by inducing expression from the *HIS3* promoter. Of the Rep constructs, only yeast co-transformed with pADRepA and pGBT9ZmRb1 grew on media lacking histidine (Table 2.3). The Rb$^-$ mutants (each with a mutation of LxCxE to LxCxK) pADRep$^{Rb\text{-}\Delta I}$, pADRep$^{Rb\text{-}\Delta C2}$, and pADRepA^{Rb-} did not interact with pGBT9ZmRb1, which is consistent with

results of Xie *et al.* (1995), and Liu *et al.* (1999a). However, yeast co-transformed with pADRep$^{\Delta C2}$ and pGBT9ZmRb1 was also unable to grow in the absence of histidine, indicating that Rep$^{\Delta C2}$ cannot interact with ZmRb1 despite having an unmutated RBR-interaction motif. This is contrary to the data of Horvath *et al.* (1998), who found that a similar-sized Rep from MSV-Ns does interact with ZmRb1. The discrepancy may reflect the differences in the assay conditions. For example, different yeast strains were used, as well as different amounts of 3-AT. The yeast strain used in this assay was more sensitive to 3-AT than the one used by Horvath *et al.* (1998), possibly resulting in this assay being too stringent (D. McGivern, pers. comm.). The different strains of MSV used could also play a role in the disparate results, although this is unlikely since MSV-Ns and MSV-Kom are closely related. Nonetheless, the main objective of the yeast two-hybrid assay was achieved, that is confirmation that the LxCxE to LxCxK mutation did indeed abolish the RBR-binding ability of the Rb$^-$ mutant Reps.

TABLE 2.3 Interactions between MSV Rep protein variants and the maize retinoblastoma protein (ZmRb1).

Binding domain fusion	Activation domain fusion	Growth of Yeast (Trp$^-$ Leu$^-$, His$^-$, + 5mM 3-AT)	Growth of Yeast (Trp$^-$ Leu$^-$, + 5mM 3-AT)
pGBT9ZmRb1	pADRepA	++	++
pGBT9ZmRb1	pRepΔC2	-	++
pGBT9ZmRb1	pADRepRb$^-$ΔC2	-	++
pGBT9ZmRb1	pADRepARb$^-$	-	++
pGBT9ZmRb1	pADRepRb$^-$ΔI	-	++
pGBT9ZmRb1	pGAD424	-	++
pLamC	pGAD424	-	++
p53	pSV40	+++	++
p53	pADRepA	-	++

Yeast cells co-transformed with pGBT9ZmRb1 and pGAD424-*Rep* fusion products were grown on drop-out medium supplemented with 5mM 3-AT in the presence or absence of histidine. Only strains containing interacting fusion proteins can grow in the absence of histidine. As negative controls, pGBT9 alone (GAL4 binding domain not fused to the ZmRb1 protein) was co-transformed with each pGAD424-*Rep* plasmid, as was pLamC, which is unable to interact with either ZmRb1 or MSV Rep. All negative controls grew in the presence, but not the absence, of histidine (data not shown). As a positive control, yeast was transformed with plasmids expressing p53 (BD fusion) and SV40 T-Ag (AD fusion). This assay was repeated with the same results

2.3.2 The Effects of Transiently Expressed Rep Proteins on MSV Replication in black Mexican sweetcorn (BMS)

Figure 2.3 uses the data of a truncated Rep construct (having been co-bombarded with pKom602 into BMS) as an example of how quantitative PCR (QPCR) was used to determine the level of

viral replication achieved in the presence and absence of various Rep constructs. In this example the Rep construct ($Rep^{\Delta Rb\Delta C2}$) enhanced replication of MSV-Kom five-fold compared with MSV-Kom alone (discussed later). This procedure was followed for each of the Rep constructs presented in Figs 2.4 and 2.5, and each bombardment was repeated at least once to ensure consistency of results. Because there is a large amount of variation between bombardments (the efficiency of one bombardment can differ greatly from another), the effect of each Rep construct on viral replication is presented relative to the replication of MSV-Kom alone (100% replication). For a diagrammatic representation of the Rep constructs used in the replication assays, refer to Fig. 2.1.

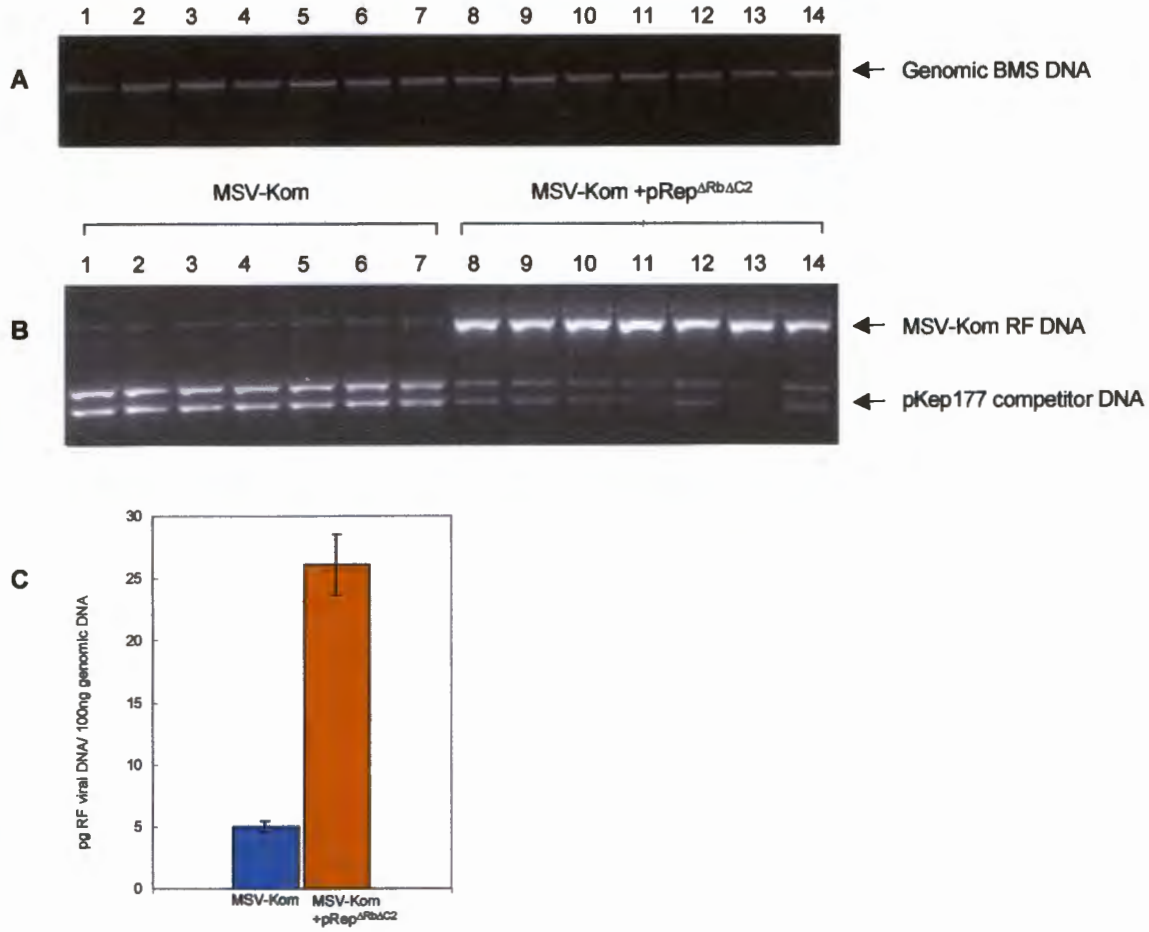

FIGURE 2.3 Determination of MSV replication levels in BMS by a quantitative PCR assay (QPCR). (A) BMS genomic DNA (extracted three days after bombardment) is quantified on a gel to ensure uniform amounts of each sample are used in the QPCR. (B) Equal amounts of genomic DNA are subjected to QPCR, each reaction being spiked with an internal control (pKep177) of known concentration. After DNA amplification, equal amounts of the QPCR reaction are digested with *Pst*1 to distinguish the spike from the viral bands, and run on a gel. (C) A densitometry programme (GelTrak) is used to determine the concentration of viral DNA in each sample, by calculating the ratio of the band intensity of viral DNA to that of the spike DNA. This is expressed as picograms (pg) of RF viral DNA present in 100 nanograms (ng) of genomic DNA.

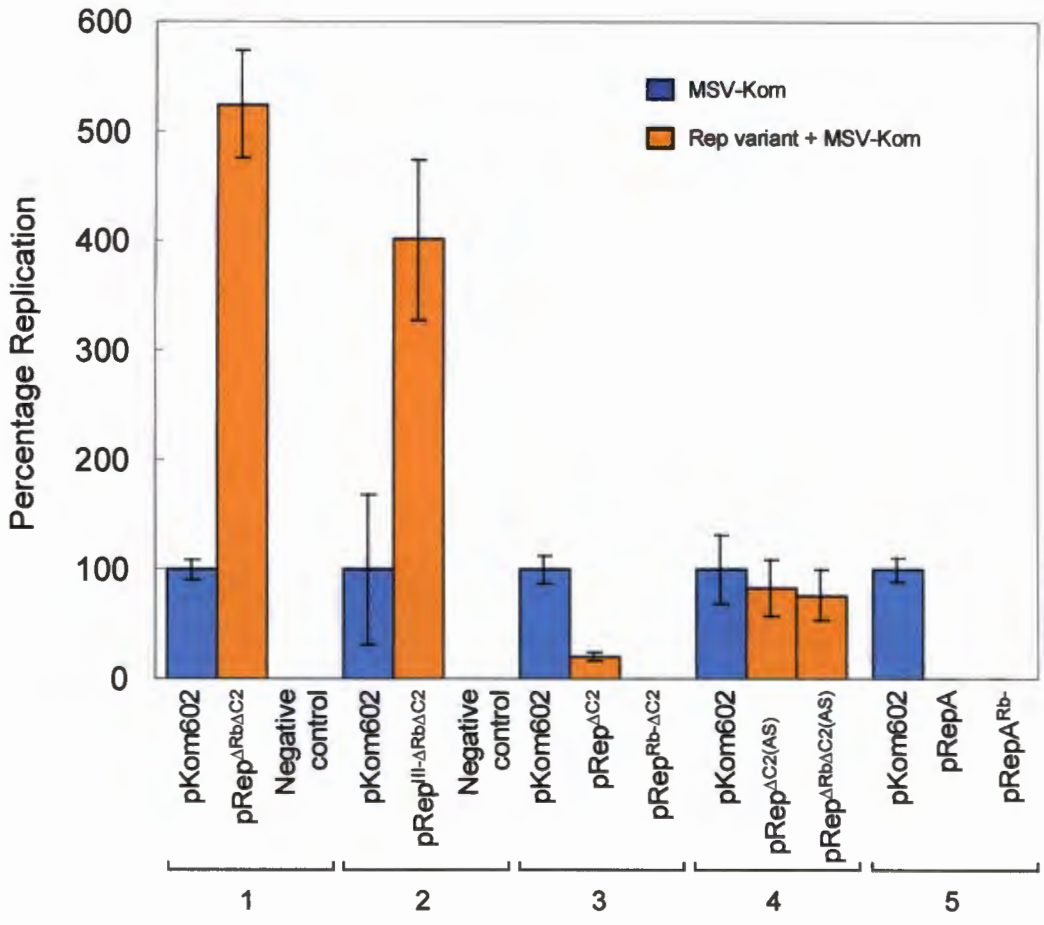

FIGURE 2.4 The effects of RepA mutants and truncations on MSV-Kom replication in black Mexican sweetcorn (BMS). In each bombardment experiment (which was repeated at least once), up to nine BMS samples were bombarded with pKom602 alone and with pKom602 + pRep$^{\Delta}$ (assays 1 and 3), pRep$^{mut\Delta}$ (assay 2), pRep$^{\Delta(AS)}$ (assay 4) and pRepA^{wt}/ pRepA^{mut} (assay 5). From individual replicate data calculated using QPCR, an average amount in pg of replicated virus in the presence and absence of each RepA derivative was calculated. The average value for MSV-Kom alone was then taken as 100% replication, and the values for MSV-Kom + each RepA derivative expressed relative to 100%. In assays 1 and 2, the negative control was genomic BMS DNA extracted immediately after bombardment with pKom602, showing that the QPCR does not amplify input plasmid pKom602. Error bars represent 95% confidence intervals.

The effects of the Rep constructs on viral replication varied enormously, from ~30-fold enhancement of replication (p*Rep* and p*Rep*$^{III+Rb-NTP+}$) to 100% inhibition (p*Rep*$^{III-Rb+NTP+}$, p*Rep*$^{III-Rb-NTP+}$, p*Rep*$^{Rb-\Delta C2}$, p*RepA* and p*RepA*$^{Rb-}$). Still others enhanced (p*Rep*$^{\Delta RbC2}$ and p*Rep*$^{III-\Delta RbC2}$) or inhibited (p*Rep*$^{\Delta C2}$) replication to a lesser extent, while the truncated antisense Reps (p*Rep*$^{\Delta C2(AS)}$ and p*Rep*$^{\Delta RbC2(AS)}$) had no significant effect. The Rb^{-} mutation, while abolishing RBR protein interaction, had no effect on the *trans*-replicating ability of pRep$^{III+Rb-NTP+}$, which enhanced the

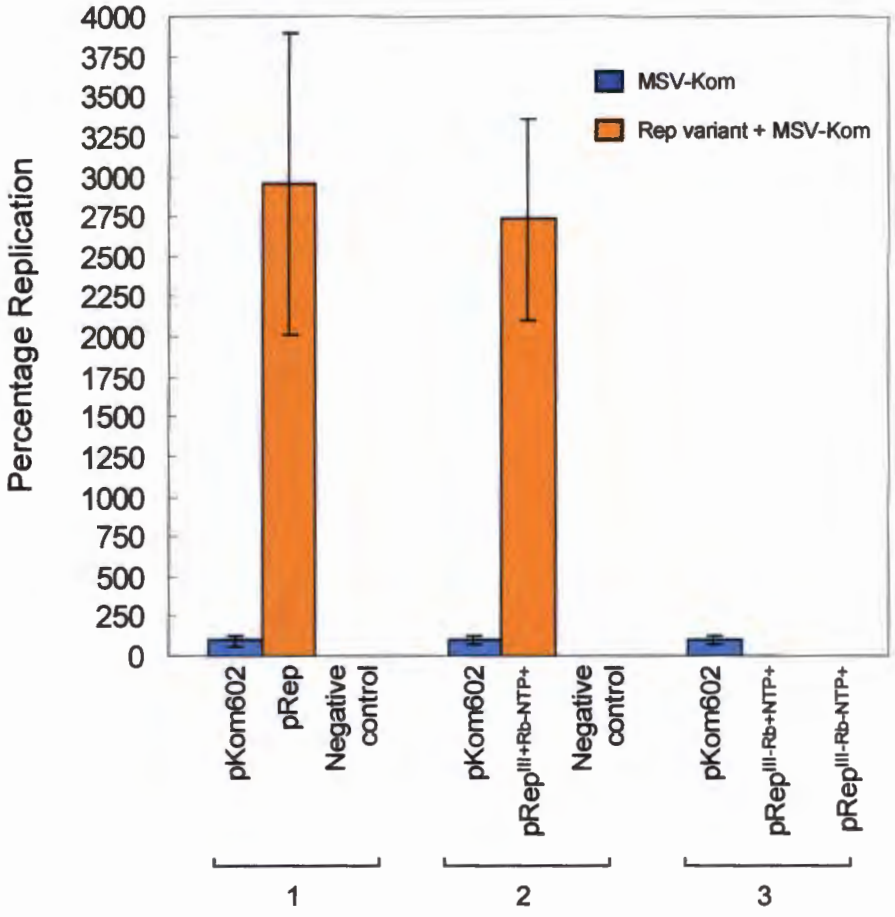

FIGURE 2.5 The effects of full-length Rep (wild-type and mutants) on MSV-Kom replication in black Mexican sweetcorn (BMS). For assays 1 and 2, the same procedure as described in Fig 2.4 was followed. Error bars represent 95% confidence intervals. Assay 3 was carried out by T. Mangwende (2000), using a Southern blot and subsequent densitometry analysis in place of QPCR.

virus' replication to the same levels as did wt Rep. A striking correlation can be seen between the size of the truncated Reps and their effects on MSV-Kom replication. The two Reps of 179 aa had the effect of enhancing replication when supplied in *trans*, while the opposite was true of the slightly larger Reps of 219 aa, which inhibited replication (see Fig. 2.1 for a diagram of these truncated Reps). The motif III$^-$ mutation in p*Rep*$^{\text{III-}\Delta\text{Rb}\Delta\text{C2}}$ had little effect on the replication-enhancing abilities of the protein, there being only a slightly significant difference in the level of viral replication achieved in the presence of p*Rep*$^{\text{III-}\Delta\text{Rb}\Delta\text{C2}}$ and the non-mutated p*Rep*$^{\Delta\text{Rb}\Delta\text{C2}}$. The Rb$^-$ mutation in p*Rep*$^{\text{Rb-}\Delta\text{C2}}$ appeared to have a more significant effect on viral replication than the non-mutated p*Rep*$^{\Delta\text{C2}}$, the former inhibiting replication completely, the latter by 80%. However, this is unlikely to be due to the abolition of RBR protein interaction, since the yeast-two hybrid

assay demonstrated that RBR protein binds to neither p*Rep*$^{\Delta C2}$ nor p*Rep*$^{Rb\text{-}\Delta C2}$ (or at the very least the RBR protein-p*Rep*$^{\Delta C2}$ interaction was so weak that it was beneath the detection levels of the assay). It is possible that the LxCxE to LxCxK mutation had another effect on the function/s of Rep apart from abolishing RBR interaction. This is the subject of Chapters 3 and 4, and will not be discussed further here.

It is likely that the Rep constructs with inhibitory effects on viral replication are behaving in a dominant negative manner. While pRep and pRep$^{III+Rb\text{-}NTP+}$ replicated MSV-Kom to very high levels, the addition of the III$^-$ mutation to both Reps led to the complete inhibition of MSV-Kom replication. Since the mutant Reps should be expressed at high levels from the ubiquitin promoter (transient expression levels, however, were not determined), they probably interfere with the functions of MSV-Kom's wt Rep (which is expressed at low levels) in various possible ways. These include the competitive occupation of viral DNA binding sites by the mutant Reps, which, once bound to the LIR, cannot initiate RCR due to the III$^-$ mutation. Although this was not directly tested on the III$^-$ mutants, the fact that pKom602 containing the III$^-$ mutation cannot replicate in BMS or establish an infection in maize (Mangwende, 2001) suggests that the mutation has abolished the ability of Rep to initiate replication. Another possibility for replication interference is the binding of the mutant Reps to wt Rep, disrupting the functions of oligomerization complexes. This could be the case with the 219 aa truncated Reps, which contain the oligomerization domain (see Fig. 2.1) and inhibit replication. Conversely, the 179 aa truncated Reps are missing the oligomerization domain (see Fig 2.1) and enhance replication. The fact that the antisense versions of both the longer and the shorter truncated Rep genes had no effect on viral replication, suggests that the inhibition and enhancement effects of the sense versions are as a result of protein expression. While oligomerization of the 219 aa Reps with MSV-Kom Rep could account for the replication inhibitory effects of the truncated proteins, the enhancement of viral replication by the 179 aa Reps is more difficult to explain. This is further explored in Chapter 5, since it is not relevant to this discussion. The inhibition of viral replication by RepA and RepA$^{Rb\text{-}}$ is also expanded on in Chapter 4, which attempts to further define the role of RepA in the virus' life cycle.

2.3.3 Transformation of *D. sanguinalis*

Before transforming *D. sanguinalis* with viral replication-interfering Rep constructs, it was important to establish the optimum pressure of the bombardment. The GUS gene (in pDPG208

under the control of the CaMV 35S promoter) was bombarded into *D. sanguinalis* calli at a pressure of 900psi and 1100psi, and three days later GUS expression was determined using a spectrofluorometric assay. In this way, a pressure of 900psi was determined to be optimum (data not shown).

From the results of the transient replication assays, p*Rep*$^{III\text{-}Rb+NTP+}$, p*Rep*$^{III\text{-}Rb\text{-}NTP+}$, p*Rep*$^{III+Rb+NTP+(AS)}$ and p*Rep*$^{Rb\text{-}\Delta C2}$ were chosen to transform *D. sanguinalis* for their ability to inhibit MSV replication (see Table 2.4 for a summary of the properties of the transgenic lines obtained). In the first two bombardments, p*Rep*$^{III\text{-}Rb+NTP+}$ and p*Rep*$^{III\text{-}Rb\text{-}NTP+}$ were each co-bombarded with pPHP7503, which contains the bar gene under the control of the CaMV 35S promoter. However, after assaying GUS expression from pAHC25 (ubiquitin promoter) compared with that from pDPG208 (CaMV 35S promoter), it was concluded that genes under the control of the ubiquitin promoter rather than the CaMV 35S promoter were more efficiently expressed in *D. sanuinalis* (data not shown). For the remainder of the *D. sanguinalis* transformations, pAHC25 (Ubi-Bar/ Ubi-Gus) was co-bombarded with p*Rep*$^{III\text{-}Rb\text{-}NTP+}$, p*Rep*$^{Rb\text{-}\Delta C2}$ or p*Rep*$^{III+Rb+NTP+(AS)}$, with the exception of pAHC*Rep*$^{III\text{-}Rb\text{-}NTP+}$ (Ubi-Bar/ Ubi-Rep) which was bombarded alone.

Having bombarded *D. sanguinalis* with p*Rep*$^{III\text{-}Rb+NTP+}$, it was speculated that the non-mutated RBR-interaction motif in the transgene could interfere with plant regeneration due to interaction with the plant RBR protein, which is important in the control of the cell cycle as well as plant differentiation and development (Huntley *et al.*, 1998). For this reason, after selecting the bombarded calli for six weeks on bialophos-containing media (MS-Bi), each bialophos-resistant callus piece (which had arisen from one transformation event) was divided into two, and one half plated onto regeneration media, while the other half was replica-plated and maintained on MS-Bi.

Some of the calli on regeneration media formed leaves and shoots, but none grew into plants, and all eventually died. After 5 months of maintenance on MS-Bi, the replica-plated calli were tested for the presence of the transgene. Out of 27 callus pieces (each an independent line), 22 were positive for *Rep*$^{III\text{-}Rb+NTP+}$ and all were positive for the bar gene. Although five callus pieces contained only the bar gene which should not interfere with regeneration, the number of full-grown plants regenerated from transgenic calli is usually far from 100% efficient even when the transgene is not toxic to growth (as was found with calli transformed with pAHC25). Out of the 22 callus pieces transgenic for *Rep*$^{III\text{-}Rb+NTP+}$, however, one would expect some regeneration of

TABLE 2.4 Properties of transgenic *D. sanguinalis*

[1]Plasmids bombarded	[2]No. of transgenic lines regenerated/ no. of calli bombarded		Comments
	Rep+Bar	Bar	
p*Rep*$^{III\text{-}Rb+NTP+}$ + pPHP7503	0/ 450 (0%)	0/ 450 (0%)	Regeneration was inhibited in callus pieces shown by PCR to be transgenic for p*Rep*$^{III\text{-}Rb+NTP+}$
p*Rep*$^{III\text{-}Rb\text{-}NTP+}$ +pPHP7503	4 /450 (0.9%)	0/450 (0%)	All lines transgenic for p*Rep*$^{III\text{-}Rb\text{-}NTP+}$ were stunted and infertile.
p*Rep*$^{III\text{-}Rb\text{-}NTP+}$ + pAHC25	1/ 450 (0.2%)	5/ 450 (1.1%)	Out of 6 regenerated lines, only one was transgenic for *Rep+bar*. This line was infertile and stunted in comparison with the *bar* transgenic lines
pAHCRep$^{III\text{-}Rb\text{-}NTP+}$	12/ 900 (1.3%)	N/A	Most (92%) transgenic lines were stunted and infertile
p*Rep*$^{III+Rb+NTP+(AS)}$ + pAHC25	0/ 450 (0%)	3/450 (0.7%)	Only plants transgenic for the bar gene were obtained.
p*Rep*$^{Rb\text{-}\Delta C2}$ + pAHC25	20/ 900 (2.2%)	2/ 900 (0.2%)	Good transformation and regeneration efficiency; a large percentage (70%) of lines transgenic for p*Rep*$^{Rb\text{-}\Delta C2}$ were phenotypically normal and fertile
pAHC25	N/A	40/ 900 (4.4%)	Very good transformation and regeneration efficiency; a large percentage of lines were phenotypically normal and fertile

[1]pPHP7503 contains the bar gene under the control of the CaMV 35S promoter; pAHC25 contains the bar gene under the control of the maize ubiquitin promoter

[2]In cases where the *Rep* construct was co-bombarded with a separate *bar* construct, not all transgenic plants contained both bar and Rep genes. Thus, the regenerated plants from each co-bombardment have been separated into those transgenic for *Rep+bar*, and those transgenic for *bar* alone. Because *bar* was the selectable marker, there were no cases of a plant being transgenic for *Rep* alone. All plants transgenic for pAHCRep$^{III\text{-}Rb\text{-}NTP+}$ contained both Rep and bar genes since they were present on the same bombarded plasmid. The presence of each transgene was determined by PCR using *Rep*- or *bar*- specific primers.

plants. Thus, it is possible that the *Rep*$^{III\text{-}Rb+NTP+}$ gene inhibited regeneration, most likely through interaction with RBR via the LxCxE motif.

Lending support to this theory is the fact that transformants of *D. sanguinalis* containing the double mutant *Rep*$^{III\text{-}Rb\text{-}NTP+}$, which cannot interact with RBR, did regenerate. However, the transformation efficiency of all bombardments with this construct was low, and as with the III^{-} mutant Rep, the double mutant Rep variant appeared to have a negative effect on aspects of plant growth. Most lines transgenic for p*Rep*$^{III\text{-}Rb\text{-}NTP+}$ were severely stunted in terms of leaf size and

height (see Fig. 2.6), and all were infertile. At the same time, the identical source calli transformed with pAHC25 alone produced phenotypically normal and fertile transgenic plants, and the transformation efficiency was high. This indicates that the stunting and infertility of plants transgenic for p*Rep*$^{III\text{-}Rb\text{-}NTP+}$ were caused by the transgene, and were not due to any negative properties of the callus from which the plants were regenerated. Although the LxCxE motif in the RBR-interaction domain was rendered non-functional in p*Rep*$^{III\text{-}Rb\text{-}NTP+}$, a second motif that may affect the normal functioning of plant cells is found within the NTP-binding domain of the C2. This motif, which contains amino acids that are conserved in the N-terminal DNA-binding domain of the avian myeloblastosis (*myb*) oncogene-homologous regulatory genes, has transcriptional activation activity (Horvath *et al.*, 1998) and it is possible that it was the cause of the stunting and infertility phenotypes. In addition, the C-terminus of the mastrevirus RepA has been shown to interact with plant GRAB (for geminivirus RepA binding) proteins, which are involved in a variety of processes, ranging from lateral root formation to development and senescence (Xie *et al.*, 1999). Interaction with GRAB proteins could therefore also interfere with plant development.

With this in mind, two truncated Rep genes were made of different sizes, both missing the entire C2 ORF and the C terminus of the RepA gene (see Fig. 2.1). As can be seen in Fig. 2.4, p*Rep*$^{Rb\text{-}\Delta C2}$ completely inhibited viral replication in BMS. Since this construct cannot interact with RBR and is missing the putative GRAB interaction domain as well as the entire C2 ORF, it was considered the best option for transformation of *D. sanguinalis*. Indeed, a large percentage of plants transgenic for p*Rep*$^{Rb\text{-}\Delta C2}$ regenerated into phenotypically normal, fertile adults. It was also notable that while many plants regenerated from calli bombarded with p*Rep*$^{III\text{-}Rb\text{-}NTP+}$ + pAHC25 and pRep$^{III+Rb+NTP+(AS)}$ + pAHC25 contained only the bar gene (indicating that bar transgenics had a selective advantage over Rep+Bar transgenics), 20 out of 22 transgenic lines regenerated from calli bombarded with p*Rep*$^{Rb\text{-}\Delta C2}$ + pAHC25 contained both Rep and bar genes.

Figure 2.6 shows three representative *D. sanguinalis* lines transformed with p*Rep*$^{Rb\text{-}\Delta C2}$ (A), p*Rep*$^{III\text{-}Rb\text{-}NTP+}$ (B) and pAHC25 (C), illustrating the effects of the different transgenes on plant growth and development. Note that (B) is very stunted in comparison with A and C. All Lines of B were infertile.

Figure 2.6 Comparison of transgenic *D. sanguinalis* lines of p*Rep*$^{Rb\text{-}\Delta C2}$ (A), p*Rep*$^{III\text{-}Rb\text{-}NTP+}$ (B) and pAHC25 (C). The plants photographed were typical of most of the plants transgenic for each *Rep* or *bar* construct.

2.3.4 Transformation of Maize Hi-II

Judging by the effects of various Rep transgenes on *D. sanguinalis* development and growth, p*Rep*$^{Rb\text{-}\Delta C2}$ was chosen to transform maize Hi-II, with pAHC25 carrying the selectable marker. 17 lines transgenic for both *Rep*$^{Rb\text{-}\Delta C2}$ and pAHC25 were regenerated. These were sent to Pannar Ltd for controlled self-pollination and setting of seed. Table 2.5 summarises the properties of the 17 lines transgenic for p*Rep*$^{Rb\text{-}\Delta C2}$, called MTA 1-17. Only five of these lines were fertile, and of those only three produced T_1 offspring. These were analysed by PCR for the presence of the *Rep*$^{Rb\text{-}\Delta C2}$ and *bar* transgenes (see Table 2.6 and Fig. 2.7). Fertile T1 plants were self-pollinated, and the number of kernels that each produced is presented in Table 2.6.

As can be seen in Table 2.5, many T_0 plants were stunted or infertile, and five of the 17 T_0 lines died while being maintained at Pannar Ltd. Although these negative effects could have been caused by the truncated Rep gene, many Hi-II plants transgenic for pAHC25 alone showed similar characteristics of stunting and infertility (data not shown). Thus, Hi-II may be more sensitive than *D. sanguinalis* to the presence of any transgene, be it *bar* or *Rep*-based. For example, integration of the transgene into an inappropriate position in the Hi-II genome could disrupt the normal functioning of important developmental genes. Also, the time spent in tissue culture, as well as the bombardment itself, could negatively affect Hi-II more than the hardier *D. sanguinalis*.

TABLE 2.5 The properties of T_0 maize lines (MTA 1-17) transgenic for *Rep*$^{Rb\text{-}\Delta C2}$ and *bar*

T_0 Line of MTA	Condition of regenerated plant	Fertility			
		No. of kernels produced by T_0 lines	No. of kernels planted	No. of T_1 offspring	No. of fertile T1 offspring
1	Good	13	10	1	1
2	Good	0	N/A	N/A	N/A
3	Good	0	N/A	N/A	N/A
4	Good	0	N/A	N/A	N/A
5	Good	1	1	0	N/A
6	Variegated	1	1	0	N/A
7	Good	142	10	10	9
8	Good	41	10	9	4
9	Died	N/A	N/A	N/A	N/A
10	Stunted	0	N/A	N/A	N/A
11	Stunted	0	N/A	N/A	N/A
12	Died	N/A	N/A	N/A	N/A
13	Died	N/A	N/A	N/A	N/A
14	Stunted	0	N/A	N/A	N/A
15	Died	N/A	N/A	N/A	N/A
16	Stunted	0	N/A	N/A	N/A
17	Died	N/A	N/A	N/A	N/A

TABLE 2.6 Transgene segregation and fertility of T_1 maize lines

T_1 Line of MTA	[1]Presence of *Rep*	[1]Presence of *Bar*	No. of kernels produced by T_1 lines
MTA1 a	✗	✗	0
MTA7 a	✓	✓	39
MTA7 b	✗	✓	16
MTA7 c	✓	✓	42
MTA7 d	✗	✓	10
MTA7 e	✓	✓	108
MTA7 f	✓	✗	13
MTA7 g	✓	✓	87
MTA7 h	✓	✓	1
MTA7 I	✗	✗	1
MTA8 a	? (See Fig. 2.7)	✗	104
MTA8 b	? (See Fig. 2.7)	✗	52
MTA8 c	✓	✗	0
MTA8 d	✓	✓	37
MTA8 e	✓	✓	0
MTA8 f	✗	✗	0
MTA8 g	✓	✓	61
MTA8 h	✓	✓	0
MTA8 I	✓	✓	0

[1] ✗ = absence of transgene
✓ = presence of transgene

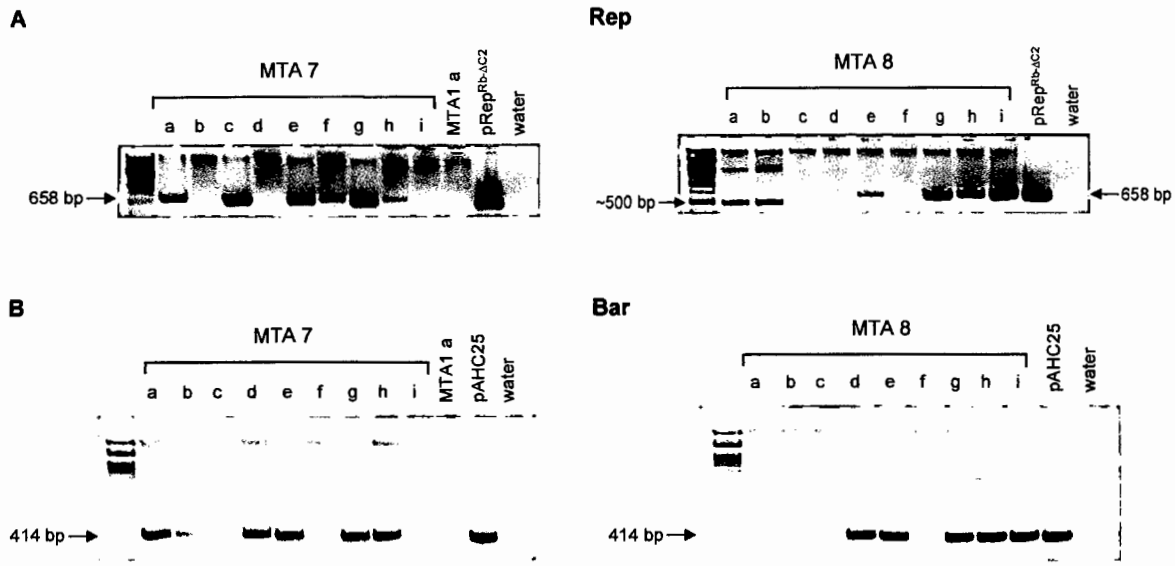

Figure 2.7 Inheritance of *Rep*$^{Rb\text{-}\Delta C2}$ and *bar* transgenes in T_1 maize. T_1 maize lines were analysed for the presence of *Rep*$^{Rb\text{-}\Delta C2}$ and *bar* transgenes by PCR using RepΔC2-specific primers (658 bp product, A) and Bar-specific primers (414 bp product, B). MTA1 a, MTA7 a-i, and MTA8 a-i are the offspring of the T_0 MTA1, 7 and 8 lines. The first lane of each gel contains λ DNA digested with *Pst*I as a size marker. The plasmids p*Rep*$^{Rb\text{-}\Delta C2}$ (A) and pAHC25 (B) are positive controls. Note the ~500 bp band for MTA8 a and b, which may be a non-specific product, or may be an amplification product of a transgene with an internal deletion.

Of the nine T_1 MTA7 lines, six contained the *Rep*$^{Rb\text{-}\Delta C2}$ gene, and seven contained the bar gene. Five out of the nine lines were positive for both *Rep*$^{Rb\text{-}\Delta C2}$ and bar genes. Similarly, of the nine T_1 MTA 8 lines, six contained the *Rep*$^{Rb\text{-}\Delta C2}$ gene, and five contained the bar gene. As with MTA7 lines, five of the nine MTA8 T_1 lines were positive for both *Rep*$^{Rb\text{-}\Delta C2}$ and bar genes. These ratios indicate Mendelian inheritance. However, it must be noted that a ~500 bp product was amplified from MTA 8 a and b lines (which were not taken as being *Rep*$^{Rb\text{-}\Delta C2}$-positive) using RepΔC2-specific primers, which could indicate the presence of a transgene with an internal deletion. This can be confirmed by sequencing the PCR product.

The fact that some lines inherited the *Rep*$^{Rb\text{-}\Delta C2}$ but not the bar gene, and vice versa, indicates that the co-bombarded plasmids did not integrate at the same loci. This could be useful when developing transgenic maize for commercial purposes, since lines transgenic for *Rep* but not *bar* could be selected, thus removing the undesirable bialophos resistance gene marker from the genome.

Two T_0 lines, MTA7 and MTA8, produced fertile T_1 offspring. The seed from nine MTA7 T_1 lines and four MTA8 T_1 lines will be tested for resistance to MSV by *Agrobacterium*-mediated delivery of a variety of MSV strains (agroinoculation; Grimsley *et al.*, 1987) into three-day old T_2 seedlings. The principle of agroinoculation is explained in Chapter 3. Before challenging maize with MSV, however, it was considered important to first test selected transgenic *D. sanguinalis* lines for resistance to MSV.

2.3.5 Challenge of *D. sanguinalis* Transgenic Plants with MSV

Due to the large number of *D. sanguinalis* lines transgenic for p*Rep*$^{III\text{-}Rb\text{-}NTP+}$ (a total of 17) and p*Rep*$^{Rb\text{-}\Delta C2}$ (a total of 20), a few selected lines (chosen for their robustness compared with other lines) were challenged with MSV. These were A6A1, A6B14, A6B15, A6B16, A6C8 and A6D10 (A6 lines, all transgenic for p*Rep*$^{III\text{-}Rb\text{-}NTP+}$) and TA1, TB1, and TB2 (T lines, all transgenic for p*Rep*$^{Rb\text{-}\Delta C2}$). Because all A6 lines were infertile, plants were regenerated from callus initiated from the parent transgenic lines, as an alternative to T_1 offspring. Callus was also initiated from the T lines even though they produced viable seed, for a number of reasons: (1) it was useful to keep a stock of each line as callus; (2) it was simple to regenerate many plants of each line from the stock callus, all of which would be genetically identical to the parental line; (3) the challenge of A6 and T lines would be more comparable if the plants came from the same source (i.e. callus instead of seed) and (4) the seed from the T lines could be kept for future work if the line looked promisingly resistant to MSV.

The first two challenges (challenge experiments 1 and 2) with MSV, carried out on lines A6A1, A6B14, A6B15 and A6B16 by exposure to viruliferous leafhoppers carrying MSV, were subsequently considered to be flawed (explained below). However, because the parental A6B14, A6B15 and A6B16 lines were eventually lost (due to fungal contamination of the stock callus) and therefore they could no longer take part in any further trials, the results from these two experiments are summarised in Table 2.8.

In the first challenge experiment, the parental T_0 plants (regenerated from bombarded callus) A6A1, A6B14, A6B15 and A6B16 were placed in a gauze-covered wooden cage, along with a *D. sanguinalis* plant transgenic for pAHC25, and a non-transgenic maize plant (cv Jubilee) as controls. ~ 20 viruliferous leafhoppers were introduced into the cage, and every second day were

TABLE 2.8 Results from two experiments challenging lines transgenic for p*Rep*$^{III\text{-}Rb\text{-}NTP+}$

Experiment	[1]Plant Lines	[2]Symptoms	[3]Presence of viral DNA
1	A6A1	-	-
1	A6B14	-	-
1	A6B15	-	-
1	A6B16	-	-
1	*D. sanguinalis* control	+++	+++
1	Maize control	+++	+++
2	A6A1 I	+++	+++
2	A6A1 II	+++	+++
2	A6A1 III	+++	+++
2	A6A1 IV	+++	+++
2	A6A1 V	-	+
2	A6A1 VI	+++	+++
2	A6A1 VII	-	+
2	A6A1 VIII	+++	+++
2	A6A1 IX	+++	+++
2	A6A1 X	+++	+++
2	*D. sanguinalis* control I	+++	+++
2	*D. sanguinalis* control II	+++	+++

[1] Lines of A6A1, A6B14, A6B15 and A6B16 were transgenic for p*Rep*$^{III\text{-}Rb\text{-}NTP+}$. In experiment 1, parental lines were used. In experiment 2, plants regenerated from callus that had been initiated from line A6A1 were used. In both experiments, the *D. sanguinalis* control was transgenic for pAHC25.

[2] Symptoms were scored four weeks after the start of the trial, and rated using the following scale: - = no streak; + = mild streak and recovery; ++ = moderate stippled streak to continuous streak; +++ = severe continuous streak.

[3] The presence of viral DNA was analysed by PCR four weeks after the start of the trial. + = very faint band (barely detectable); ++ = faint, but easily detectable band; +++ = very bright band.

shaken off the plants and allowed to resettle to ensure that most plants would be fed on. After four weeks, DNA was extracted from each of the plants' leaves and analysed for the presence of MSV by PCR. The same procedure was followed for the second challenge experiment, this time using 10 plants regenerated from stock callus initiated from the A6A1 parental line, and two control plants transgenic for pAHC25. As can be seen in Table 2.8, some of the *Rep* transgenic plants did not become infected with MSV while all the controls did. In particular, all four parental lines appeared resistant, although only two out of ten plants regenerated from the A6A1 callus were unsusceptible to MSV. However, it could not be discounted that the uninfected plants were not fed on by the leafhoppers, or at least for long enough for viral transmission to occur. The time taken for the leafhopper's proboscis to reach the phloem and transmit the virus can take as long as three hours from initial access (Bosque-Pérez, 2000). Because the *Rep* transgenic plants' leaves were very small and perhaps not very appetising to the leafhoppers in comparison to the control plants (see Fig 2.6), the leafhoppers may have preferentially fed on the control plants, skewing the results. Therefore, in subsequent challenge experiments, the viruliferous leafhoppers were placed in a small vial containing a ~1 mm slit, through which a single leaf from a transgenic plant

was inserted, ensuring that the leafhoppers in a particular vial could only feed on one leaf of one plant. Three vials were placed at different positions on each plant (Fig. 2.8A). In total the same number of leafhoppers were fed on each plant, as far as possible ensuring uniformity of the MSV inoculum received by the plant.

A preliminary challenge employing this method was carried out at Pannar Ltd by Dr. R. Kloppers, using five plants of line A6A1 and three plants of line A6C8, as well as a field-grown non-transgenic *D. sanguinalis* control (challenge no. 3). Whereas the non-transgenic control plant developed symptoms 10 days after the start of the challenge, no symptoms were observed on any of the transgenic plants throughout the two-month trial (Fig.2.8B and C). Viral DNA levels in the challenged transgenic plants were shown by PCR to be either non-existent or significantly lower than those in the non-transgenic control plant (Fig. 2.8D). The A6C8 line appeared to be particularly resistant. Not only did the plants develop no symptoms, but viral replication was also completely inhibited in two and significantly reduced in one out of the three challenged A6C8 plants. Viral replication in three out of five A6A1 plants was greatly reduced compared with that in the non-transgenic control plant, and two out of the five A6A1 plants contained no detectable viral DNA.

In a fourth challenge, six plants of line TB1 and one plant of line TB2, both transgenic for p*Rep*$^{Rb\text{-}\Delta C2}$, and seven control plants transgenic for pAHC25 were challenged with MSV (Table 2.9). Leafhoppers in vials were fed first on the control plants for two days and then transferred to the plants of line TB1 and TB2 and allowed to feed for two days. The plants were monitored for symptom development for five weeks. The same procedure was followed for challenges five (four control plants and four plants of line TB1) and six (one control plant and one plant of line A6C8). Symptoms and viral DNA amplified by PCR from total DNA extracted from challenged plants' leaves were scored as in Table 2.8.

In the experiments shown in Table 2.9, each *bar* transgenic control plant can be directly compared with its *Rep*-transgenic counterpart (e.g. control 1 with TB1 I; control 2 with TB1 II, etc), since the same leafhoppers, transferred from the control to the corresponding *Rep* transgenic, fed on both plants. Leafhoppers remain viruliferous for their lifetime, and therefore should transmit the virus throughout the four days they feed on the plants (two days on the *bar* controls followed by two days on the *Rep* transgenics). It can be seen that in experiment 4, *bar* controls 1,

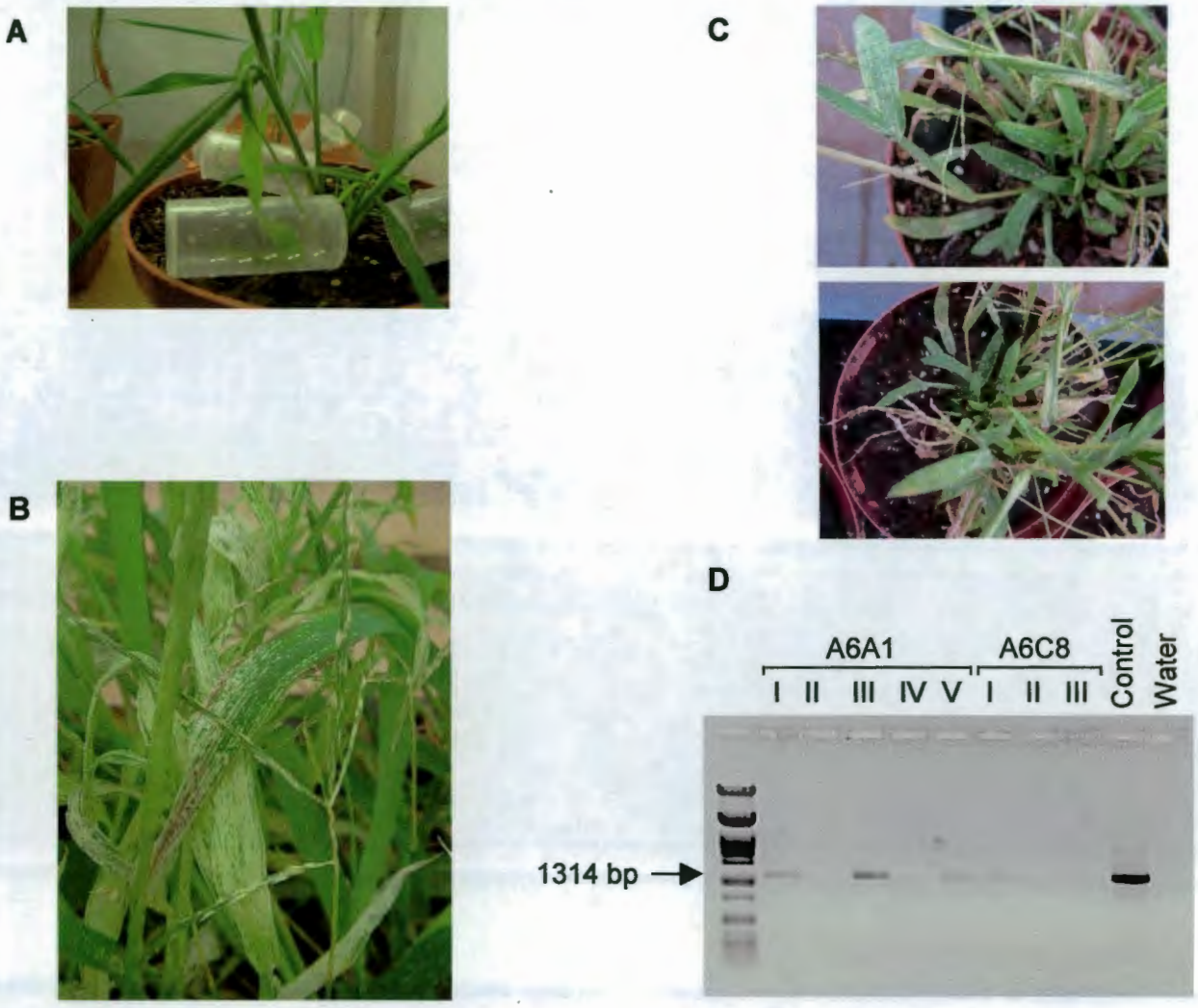

FIGURE 2.8 Challenge with MSV of lines A6A1 and A6C8, both transgenic for p*Rep*$^{III\text{-}Rb\text{-}NTP+}$. Four leafhoppers were placed in each vial, and three vials placed at different positions on each plant (A). The *D. sanguinalis* control became severely infected (B) while the *Rep* transgenics were asymptomatic (C, showing two lines of A6A1). PCR analysis of viral replication in leaves of the transgenic and non-transgenic *D. sanguinalis* (D) showed greatly reduced levels of viral replication in the transgenics' leaves compared with the control. The degenerate primers MSVDEG1 and MSVDEG2 (described in section 2.2.3.3) were used to amplify the putatively mixed MSV population transmitted by the leafhoppers. These primers were designed to amplify DNA fragments from all African streak viruses sequenced up until 1996 (Willment *et al.*, 2001). Note in (C) that the leaves of A6A1 plants are stunted in comparison with the control in (B). Although faint speckles can be seen on some of the transgenics' leaves, these looked more like leafhopper–feeding damage rather than symptoms. This was confirmed by the PCR results.

3, and 6 developed symptoms 13 days after the start of the challenge, while the corresponding TB1 I, III, and VI plants never developed an infection.

TABLE 2.9 Results of challenge experiments 4, 5 and 6

Challenge experiment number	*Bar* transgenic plants (controls)	Time to develop symptoms (Days)	[1]Symptom severity after 35 days	[2]Presence of viral DNA	[3]*Rep* transgenic plants	Time to develop symptoms (Days)	[1]Symptom severity after 35 days	[2]Presence of viral DNA
4	1	13	+	-	TB1 I	-	-	-
	2	12	+++	++	TB1 II	14	+	+
	3	13	++	+	TB1 III	-	-	-
	4	12	+	+	TB1 IV	14	+	-
	5	12	+	+	TB1 V	14	+	-
	6	13	+++	+++	TB1 VI	-	-	-
	7	13	+++	+++	TB2 I	14	+++	+++
5	1	11	+++	ND	TB1 VII	15	++	ND
	2	-	-	ND	TBI VIII	-	-	ND
	3	11	+++	ND	TB1 IX	11	++	ND
	4	11	+++	ND	TB1 X	11	+++	ND
6	1	11	+++	+++	A6C8 VI	-	-	-

[1]Symptoms were scored five weeks after the start of the trial, and rated using the following scale: - = no streak; + = mild streak and recovery; ++ = moderate stippled streak to continuous streak; +++ = severe continuous streak.

[2]The presence of viral DNA was analysed by PCR five weeks after the start of the trial. + = very faint band (barely detectable); ++ = faint, but easily detectable band; +++ = very bright band. ND = Not determined

[3]TB1 and TB2 plants are transgenic for p*Rep*$^{Rb\text{-}\Delta C2}$; A6C8 is transgenic for p*Rep*$^{III\text{-}Rb\text{-}NTP+}$

Others, such as TB1 II, did become infected after 14 days, but showed a recovery phenotype (+) and correspondingly a low level of viral DNA was amplified from this plant's leaves. Conversely control 2, which became infected after 12 days, had a severe infection (+++), and high levels of viral DNA in its leaves. The remainder of the TB1 plants in challenge 4 did not differ greatly in their response to viral infection from their control counterparts. The line TB2 was highly susceptible to virus infection, as was its corresponding control.

The TB1 plants in challenge experiment no. 5 showed less resistance to viral infection than those in challenge 4, all but one (TB1 VIII) becoming infected. The fact that TB1 VIII was not symptomatic by day 35 was probably due to weakly- or non-virulent leafhoppers, since the corresponding control plant also remained symptomless. This illustrates the importance of having the same leafhoppers feed on both control and *Rep* transgenic plants. The infection of TB1 VII was slightly delayed compared with its control, although only by four days, and symptoms were slightly less severe than those of the control. TB1 VII also developed symptoms in fewer stalks than the control. The same effects could be seen in TB1 IX, whose symptoms were slightly attenuated and limited to three stalks, compared with ten in the control, which developed a very severe infection.

In challenge no. 6, A6C8 developed no symptoms (as in challenge 3) and no viral DNA was detected in the plant's leaves, whereas the control developed a severe infection and high levels of viral DNA were amplified from its leaves by PCR.

In a seventh and final trial, five plants initiated from a line transgenic for p*Rep*$^{Rb\text{-}\Delta C2}$ (TA1) and five non-transgenic *D. sanguinalis* plants were challenged with MSV. The analysis of the symptoms was different from the previous trials, in that symptom severity was scored using a key of symptoms (see Appendix B) ranging from 1% (very mild) to 95% (very severe) chlorosis.

All control plants developed symptoms between 15 and 17 days after the start of the challenge, whereas the transgenic TA1 plants had different responses to the challenge (Table 2.9). All TA1 plants showed some kind of resistance, from immunity (TA1 V) to delayed symptom development (TA1 II) to attenuated symptoms compared with the corresponding control (TA1 I and III). The severity of the symptoms of TA1 IV equaled that of the control. However, whereas the control developed symptoms in all leaves that emerged after MSV transmission, only the leaves of one stalk of TA1 IV developed symptoms. A similar pattern could be seen in all the other TA1 plants, indicating that the spread of the virus from initially infected leaves was inhibited. The restriction of symptoms in plant TA1 I compared with the corresponding control plant is shown in Fig. 2.9A, B and C.

To analyse the viral infection at a molecular level, the DNA of three leaves taken from each plant 47 days after the start of the trial was subjected to PCR (Fig. 2.9D). From the TA1 plants showing symptoms, a young (newly-emerged) symptomatic leaf (A), a young asymptomatic leaf immediately adjacent to the symptomatic leaf (B), and a young asymptomatic leaf on a different stalk distant from the symptomatic leaf (C) was analysed by PCR to determine if there was any virus spread from the symptomatic leaf. From the non-transgenic control plants three young leaves were taken from three different areas of the plant. Figure 2.9D shows that viral DNA was present only in the symptomatic leaves of the transgenic plants TA1 Ia, IIIa, IVa and all control plants. In the TA1 plants, even leaves immediately adjacent to the symptomatic leaves contained no viral DNA. TA1 V at this stage in the experiment (47 days) did not have any new symptomatic leaves - the one stalk containing an infected leaf had died and no new infection emerged, which correlates to the lack of viral DNA in any of the leaves analysed. Visible symptoms on TA1 III only emerged on one leaf on day 47, and subsequently viral DNA was amplified from this leaf (IIIA) by PCR.

TABLE 2.9 Results of challenge experiment no. 7 using five plants of line TA1, transgenic for p*Rep*$^{Rb\text{-}\Delta C2}$, and five non-transgenic control plants

Challenged Plant	Time to develop symptoms (days)	[1]Symptom severity after 35 days	Comments
Control 1	17	93.7±2.5	All new leaves subsequent to the challenge were symptomatic
Control 2	17	76.6±8	All new leaves subsequent to the challenge were symptomatic
Control 3	15	85±4.6	All new leaves subsequent to the challenge were symptomatic
Control 4	17	41.3±6.1	All new leaves subsequent to the challenge were symptomatic
Control 5	17	86.3±6.1	All new leaves subsequent to the challenge were symptomatic
TA1 I	19	59.1±5.8	Symptoms restricted to three stalks. The rest of the plant remained symptomless
TA1 II	47	-	Very mild symptoms seen only after 47 days near the end of the trial period; restricted to 1 stalk. The rest of the plant remained symptomless
TA1 III	22	36.4±11.1	Symptoms restricted to 1 stalk. The rest of the plant remained symptomless
TA1 IV	18	42.4±15.3	Symptoms restricted to 1 stalk. The rest of the plant remained symptomless
TA1 V	-	-	No symptoms developed throughout the trial

[1]average % chlorosis per leaf; 95% CI

In conclusion, the challenges of various transgenic lines with MSV produced varied although encouraging results, summarised below.

Of 16 challenged A6A1 plants in three separate challenge experiments, eight showed MSV resistance, manifesting itself in these plants as no symptom development and vastly lowered viral replication compared with controls.

The parental lines A6B14, A6B15 and A6B16 did not become infected in challenge no.1, but it could not be discounted that the "resistance" may have been due to non-transmission of MSV by the leafhoppers due to preferential feeding on the more succulent-looking control plants.

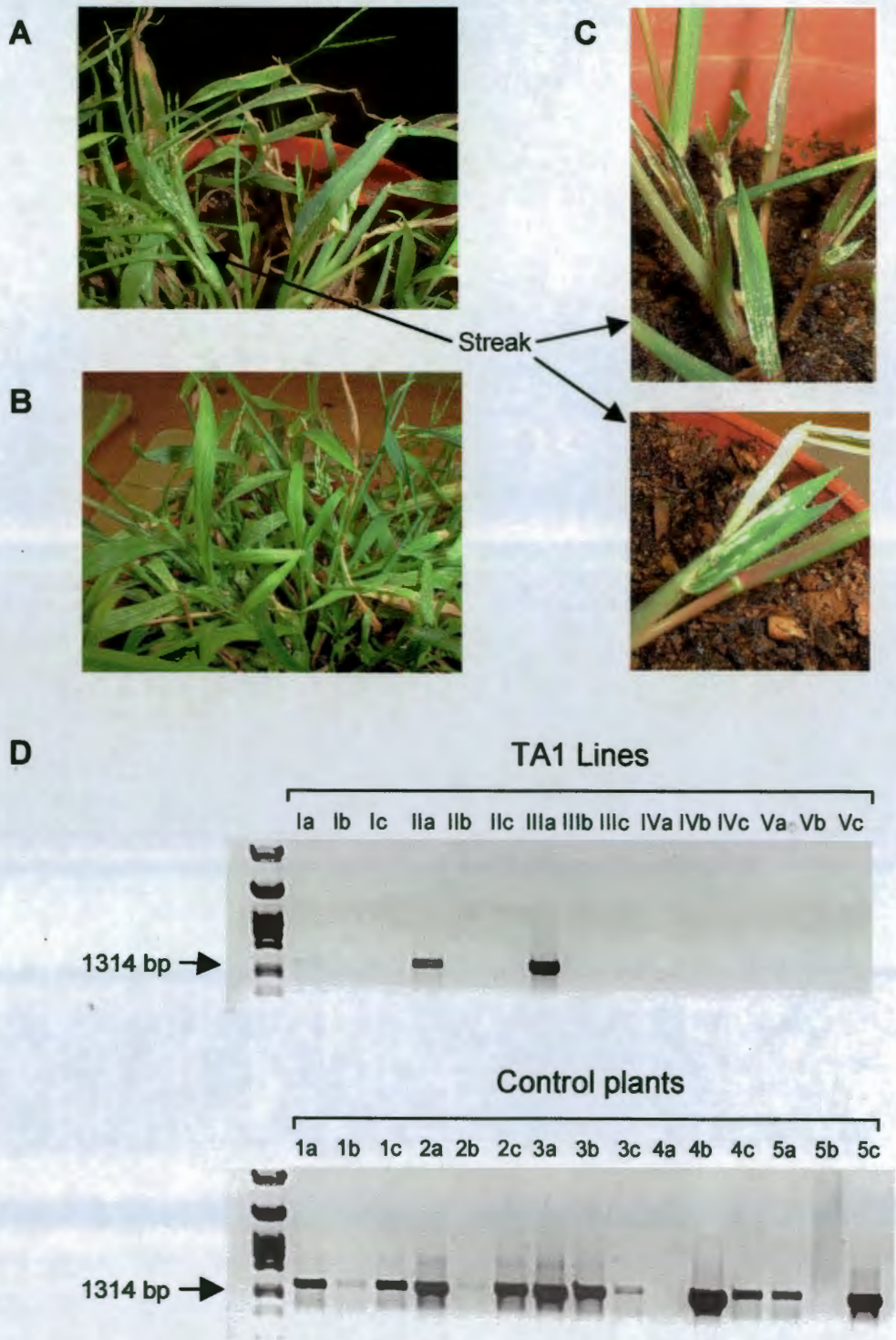

FIGURE 2.9 The challenge of line TA1, transgenic for p*Rep*$^{Rb\text{-}\Delta C2}$, with MSV. Plant TA1 I is shown in (A) and (B). In (A), symptoms on a leaf of one stalk are shown. In (B), it is clear that leaves on other stalks of the plant are symptomless. In contrast, the leaves of all stalks of the corresponding control plant (C) are severely infected, and the plant is close to dying. The photographs were taken 35 days after the start of the challenge. Viral replication in the plants' leaves was analysed by PCR (D) 47 days after the start of the challenge. For TA1 plants, a = symptomatic leaf (except in the case of non-symptomatic plants), b = adjacent non-symptomatic leaf; c= distant non-symptomatic leaf. For control plants, a, b and c are leaves taken from 3 distant areas of the plant.

Three lines transgenic for p*Rep*$^{Rb\text{-}\Delta C2}$, TB1, TB2 and TA1 had different responses to the challenge with MSV. While TB2 was highly susceptible, some plants of TB1 showed differing degrees of resistance to MSV, as seen in Table 2.9. Overall, however, TB1 does not look like a promising line in terms of reliable, consistent MSV resistance.

The p*Rep*$^{Rb\text{-}\Delta C2}$ transgene in all TA1 plants appeared to confer phenotypes of delayed, attenuated symptoms, as well as potentially restricting movement or spread of the virus. The TA1 plants that developed symptoms only did so in three or fewer stalks, which were most likely the points of entry for the virus; i.e. the stalks on which the viruliferous leafhoppers were fed. From the point of entry it is possible that the virus was not able to replicate to titres high enough to enable spread to surrounding stalks.

The most encouraging line in terms of consistent MSV resistance was A6C8. Four out of four plants in two separate challenge experiments never developed symptoms, while the controls were severely infected. In addition, only one of the four challenged plants contained any viral DNA detectable by PCR. Since the possibility that the titre of MSV in viruliferous insects may decrease when fed first on control plants (non-transformed or pAHC25-transformed) was considered, a larger-scale challenge experiment involving more A6C8 plants was attempted, in which the leafhoppers were to be fed first on the transgenic plants. Unfortunately, after four years in tissue culture, the callus appears to have lost its ability to regenerate into plants. Indeed, the reason only one A6C8 plant was used in challenge no. 6 is because it was the only one to regenerate from a plateful of callus. Thus, the question of whether the efficiency of transfer of MSV to the transgenic plants is affected by being passaged first through non-transgenic plants remains to be answered. However, bearing in mind that A6C8 plants developed no symptoms at all, it is perhaps unlikely that virus titres in leafhoppers would decrease so dramatically over a period of two days as to render the leafhoppers non-viruliferous.

Another challenge, using plants of line A6D10 (transgenic for pAHC*Rep*$^{III\text{-}Rb\text{-}NTP+}$) was aborted because the regenerated plants were too stunted to enable the attachment of leafhopper-containing vials at three separate positions on each plant, and at an early stage in their development the plants stopped growing any new leaves. This could have been due to effects of the *Rep*$^{III\text{-}Rb\text{-}NTP+}$ transgene, or to too much time spent in tissue culture as callus (or a combination of both). There are many more transgenic lines being maintained as callus, and even if they cannot regenerate plants they may still be useful in identifying transgenes that confer MSV resistance. For example,

different transgenic callus lines can be bombarded with MSV, and the levels of replication supported by the transgenic callus compared with control callus can be determined using quantitative PCR, as in the transient assays in BMS.

The differences in the degree of resistance of the challenged transgenic plants could be due to differing expression levels of the transgenes. This was determined using reverse transcription PCR, described in section 2.3.6.

2.3.6 Expression of Transgenes in *D. sanguinalis* Lines used in MSV-Challenge Experiments

The expression of GUS (in cases where the transforming plasmid contained the *uidA* gene), Bar and Rep was determined in the transgenic lines that were challenged with MSV, by histochemical staining and RT-PCR of RNA extracted from the transgenic callus lines. Figure 2.10 shows GUS stains of lines A6B14, A6B15, A6B16, A6C8, TA1, TB1 and TB2. A6A1 and A6D10 were not included since neither line was transformed with the GUS gene. Since there were no Rep antibodies available for analysis of Rep expression by western blotting, expression of GUS gives an idea of how well the transgenes are being expressed by each line, and RT-PCR of the *Rep* transcript provides indirect evidence that the protein is being expressed. As can be seen in Fig. 2.10, GUS expression was not detected in leaf material as clearly as it was in callus. In (A), the leaves of A6B15 and TA1 were not positive for GUS, while blue spots could be seen in calli of the same lines. The leaf of A6C8, although positive for GUS, also did not stain as well as the A6C8 callus. In a previous study Chen (1996) found that while GUS activity in transgenic plants' leaves was not detectable by histochemical staining, a more sensitive fluorescence assay (where the protein was extracted from the leaves) detected GUS activity in the majority of the same plants. Thus, the lines TB1 and TB2 may be expressing GUS, albeit at lower levels than A6C8, even though the leaves gave negative results. Despite the fact that genetically identical cells should comprise the calli of an individual line, expression of GUS was sporadic in all lines but A6C8. This could provide a clue as to why plants regenerated from callus had different responses to MSV infection, from sensitivity to immunity. If Rep expression in callus cells follows the same pattern as that of GUS, it is conceivable that some plants may regenerate from cells expressing Rep, while others may arise from cells not expressing Rep. The majority of A6C8 calli expressed GUS, which is consistent with the fact that 100% of plants regenerated from A8C8 calli were resistant to MSV infection, i.e. it is probable that all the plants were expressing $Rep^{III\text{-}Rb\text{-}NTP+}$. Unfortunately, leaf samples taken before the plants were challenged with MSV

were used for DNA extractions to confirm the presence of the transgene, and not for transgene expression analysis.

While A6C8 and A6D10 contain relatively high levels of Rep transcript (Fig. 2.11), cDNA bands in TA1, TB1 and TB2 samples are very faint. The Rep transcript levels of all lines correlate with the GUS expression levels shown by histochemical staining of the calli. Since the RNA was extracted from calli, it follows that calli in which most cells are expressing the transgene, such as A6C8, will collectively have higher levels of transcript, while in other lines a low percentage of callus cells may be expressing the transgene, diluting the transcript out. Only calli expressing Bar can survive on media containing bialophos (on which the callus lines were constantly maintained) explaining the higher levels of bar transcript in all lines.

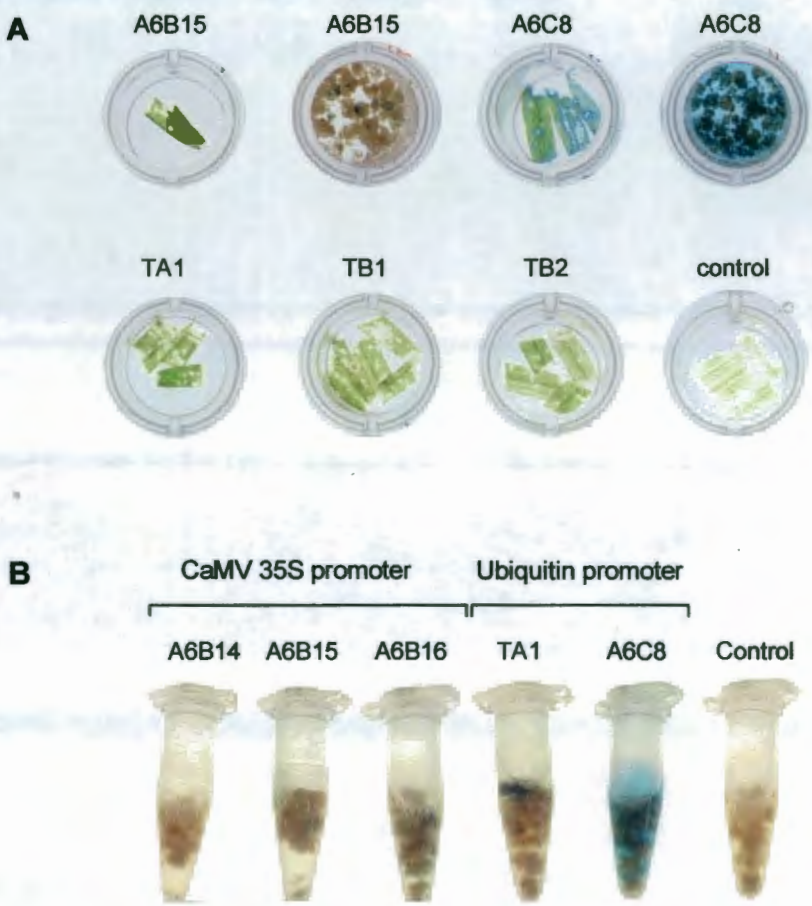

FIGURE 2.10 GUS stains of callus and leaf material of *Rep* transgenic lines. A6B14, A6B15, A6B16 and A6C8 are lines transgenic for $Rep^{III\text{-}Rb\text{-}NTP+}$. TA1, TB1 and TB2 are lines transgenic for $pRep^{Rb\text{-}\Delta C2}$. The control is leaf or callus material from a non-transgenic plant. In (A) leaf material was treated with the X-Gluc substrate and then destained in order to see blue spots more easily. Most of the leaf material did not stain blue, but blue-stained calli of two lines, A6B15 and A6C8, are shown to compare the sensitivity of the histochemical staining technique in leaves and calli. In (B) stained calli are shown. A6B14, A6B15 and A6B16 were transformed with the GUS gene under the control of the CaMV 35S promoter, while in A6C8 and TA1 GUS is expressed from the maize ubiquitin promoter.